(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 January 2002 (24.01.2002)

PCT

(10) International Publication Number
WO 02/06469 A2

(51) International Patent Classification[7]: C12N 15/10

(21) International Application Number: PCT/US01/22640

(22) International Filing Date: 18 July 2001 (18.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/218,921	18 July 2000 (18.07.2000)	US
60/219,085	18 July 2000 (18.07.2000)	US
09/692,732	19 October 2000 (19.10.2000)	US
09/691,873	19 October 2000 (19.10.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/218,921 (CIP)
Filed on	18 July 2000 (18.07.2000)
US	60/219,085 (CIP)
Filed on	18 July 2000 (18.07.2000)
US	09/691,873 (CIP)
Filed on	19 October 2000 (19.10.2000)
US	09/692,732 (CIP)
Filed on	19 October 2000 (19.10.2000)

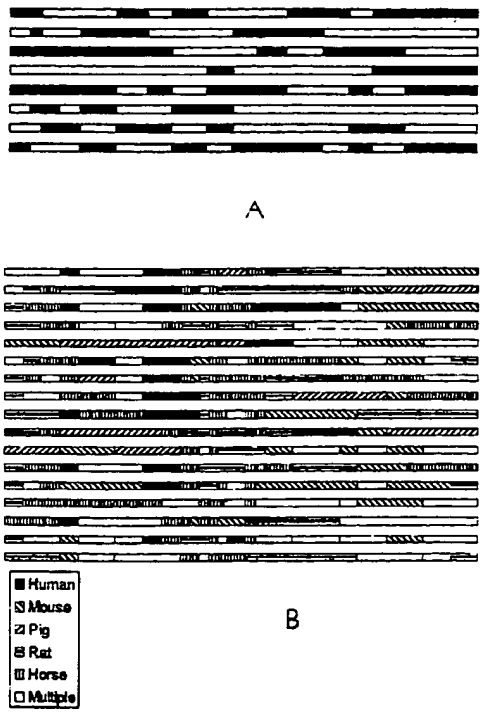
(71) Applicant *(for all designated States except US)*: **ENCHIRA BIOTECHNOLOGY CORPORATION** [US/US]; 4200 Research Forest Drive, The Woodlands, TX 77381 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **COCO, Wayne, M.** [US/US]; 43 Cornerbrook Place, The Woodlands, TX 77381 (US). **ENCELL, Lance, P.** [US/US]; 58 Brooksedge Circle, The Woodlands, TX 77382 (US). **ARENSDORF, Joseph, J.** [US/US]; 27225 Orth Lane, Oak Ridge North, TX 77385 (US).

[Continued on next page]

(54) Title: METHODS OF LIGATION MEDIATED CHIMERAGENESIS UTILIZING POPULATIONS OF SCAFFOLD AND DONOR NUCLEIC ACIDS



(57) Abstract: The present invention is drawn to a method for forming at least one chimeric polynucleotide, methods for directed evolution, chimeric polynucleotides and libraries of chimeric polynucleotides. One method comprises contacting a first population of single-stranded oligonucleotides wherein the oligonucleotides share minimal complementarity with each other with a second population of oligonucleotides, under conditions wherein the oligonucleotides of the first and second populations hybridize to each other, forming at least one hybridized complex, comprising at least one polynucleotide from the first population hybridized to at least two oligonucleotides from the second population. Single-stranded regions are filled in using polymerase. The filled-in hybridized complex is treated such that the adjacent nucleic acids are ligated, forming at least one chimeric polynucleotide.

WO 02/06469 A2

(74) **Agents:** **ELMORE, Carolyn, S.** et al.; Hamilton, Brook, Smith & Reynolds, P.C., 530 Virginia Road,, P.O Box 9133, Concord MA 01742-9133 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS OF LIGATION MEDIATED CHIMERAGENESIS UTILIZING POPULATIONS OF SCAFFOLD AND DONOR NUCLEIC ACIDS

RELATED APPLICATION

This application is a continuation-in-part of United States Application No. 09/692,732 filed October 19, 2000 and United States Application No. 09/691,873 filed October 19, 2000. This application also claims the benefit of United States Provisional Application No. 60/219,085, filed July 18, 2000 and United States Provisional Application No. 60/218,921 filed July 18, 2000. The teachings of all the above-referenced applications are hereby incorporated by reference in their entities.

BACKGROUND OF THE INVENTION

Genetic improvements occur more frequently when the generation of mutations is coupled with genetic recombination. Recombination between similar but non-identical polynucleotide targets allows for the consolidation of favorable mutations that appear on separate copies of the target, as well as the elimination of detrimental mutations (Harayama, S. *Trends Biotechnol.*, 16:76-82 (1998)). The effect of genetic recombination on the fixing of multiple beneficial mutations is noticeable when comparing sexually and asexually replicating organisms. Although single mutation rates are generally similar for sexually and asexually replicating organisms, juxtaposing these rare mutation events speeds up the evolutionary process dramatically. It is this ability to combine beneficial mutations and rapidly eliminate deleterious but not lethal mutations that enables sexually replicating organisms to evolve at a faster rate than asexually replication organisms. The reduction in evolutionary potential in asexually replicating populations is known as Müller's ratchet (Müller, H., *Mut. Res. 1*:2-9, 1964). The process of altering genetic functions through generation of mutants, and/or chimeric genetic recombinants,

coupled with selection and/or screening is termed "directed evolution."

The ability to generate a chimeric polynucleotide is fundamental to the process of directed evolution. Chimeric polynucleotides can result from recombination between two or more parent polynucleotides. To date, various strategies have been described to accomplish *in vitro* recombination. These include the following: "sexual PCR" (Stemmer, W. *Nature, 370*:389-391, 1994; United States Patent Nos. 5,605,793 and 5,811,238), which utilizes fragments cleaved from two or more parent double-stranded polynucleotides to form mutagenized double-stranded polynucleotides; "StEP" (Zhao, H. *et al., Nat. Biotechnol. 16*:258-61, 1998), which is characterized by multiple rounds of incomplete elongation of primers on variant templates; and the "RACHITT"™ method (Coco, W. *et al., Nat. Biotechnol. 19*:354-359, 2001), which typically uses the strategy of hybridizing fragments from one or more parents polynucleotides to a transient template, treating the overlaps and gaps enzymatically to yield a linear final product, and then destroying the original template prior to cloning.

These methods of directed evolution can form libraries of "chimeric" polynucleotides, so called because they include recombined sequences from more than one parent gene. The library of chimeric products, however, for sexual PCR and StEP, generally represent a limited and biased sampling of all potential chimeric products. These deficiencies are, in part, a result of limitations inherent in the essential PCR step in each of these methods. Moreover, these methods can suffer from "blind spots" in the gene or polypeptide of interest, where exchanges between parental DNA from two or more sources is rare or nonexistent due to the manner in which the DNA is fragmented or because regions of homology of a certain size are generally required to allow homologous recombination between the parental DNA. Although RACHITT™ overcomes these deficiencies to generate a broad library of chimeric polynucleotides, other distinct methods capable of producing libraries of chimeric polynucleotides of differing complexity would enhance the progress of directed evolution.

SUMMARY OF THE INVENTION

The methods of the present invention facilitate the generation of chimeric polynucleotides and do not require hybridizing donor fragments to a target- or full-length template. Rather, a first population of hybridizing donor fragments, *e.g.*, oligonucleotides, are assembled using a second population of scaffold fragments, *e.g.*, oligonucleotides, to form a double-stranded chimeric polynucleotide in which one or both strands are chimeric. One strand of the chimeric double-stranded polynucleotide can comprise, for example, scaffold fragments and regions between the scaffold fragments that were filled-in during the process; the opposite strand can also comprise donor fragments and regions between the donor fragments that were filled-in during the process. Because the chimeragenesis process of the present invention does not rely upon a contiguous, full-length template, it is unnecessary to modify a template to facilitate its removal.

In one embodiment, the invention is directed to a method for forming a chimeric polynucleotide including the steps of: contacting a population of single-stranded scaffold fragments with a population of donor fragments under conditions such that at least one scaffold fragment hybridizes to at least two donor fragments at distal regions of the scaffold fragment; treating the hybridized complexes such that single-stranded regions of the hybridized complex are filled-in; and treating the filled-in hybridized complexes such that adjacent fragments are ligated, forming a chimeric polynucleotide. In a particular embodiment, the method can also include the step of trimming flaps. Scaffold fragments can contain sequences of from about 10 to about 1000 nucleotides in length, preferably from about 25 to 100 nucleotides in length. Also, scaffold fragments can be derived from a single strand of a parent polynucleotide. Donor fragments can contain sequences of about 10 to about 1000 nucleotides in length, and they can be single-stranded. Donor fragments can be derived from a single strand of a parent polynucleotide. In a particular embodiment, scaffold and donor fragments hybridize to each other under conditions of low stringency. The population of scaffold fragments can be produced synthetically, or they can be produced by cleaving a polynucleotide of interest that is a full-length cDNA. The population of scaffold or donor fragments can include a fragment with

at least one region of random sequence. The method can further include a step of preparing at least one single-stranded population of scaffold fragments, derived from a randomly fragmented single-stranded polynucleotide of interest. In one embodiment, the populations of scaffold and donor fragments are sufficient to form a full-length chimeric polynucleotide. In a particular embodiment, the invention includes the step of screening or selecting at least one chimeric polynucleotide having desired characteristics. In another aspect, the invention is directed to chimeric polynucleotides prepared according to the methods described herein.

In another aspect, the invention is directed to a library of chimeric polynucleotides prepared according to the methods described herein. The library can be such that the majority of the chimeric polynucleotides contain at least 3 crossover sites. The library can contain at least one chimeric polynucleotide which contains the number of crossovers approaching the theoretical limit. The library can contain at least five chimeric polynucleotides which contains the number of crossovers approaching the theoretical limit.

In another embodiment, the invention is directed to a method for forming at least one double-stranded chimeric polynucleotide having desired characteristics including the steps of: contacting a population of scaffold fragments derived from a template polynucleotide with a population of donor fragments under conditions such that fragments of the scaffold and donor populations can hybridize to each other; forming at least one hybridized complex comprising at least one scaffold fragment hybridized to at least two donor fragments; treating the hybridized complex such that single-stranded regions of the hybridized complex are filled-in; treating the filled-in hybridized complex such that adjacent fragments are ligated, thereby forming a double-stranded chimeric polynucleotide. In one embodiment, the invention also includes the steps of trimming flaps and/or screening or selecting at least one double-stranded chimeric polynucleotide having desired characteristics. Scaffold fragments can contain sequences that are at least about 25 percent as long as a gene of interest. Scaffold and/or donor fragments can contain sequences of from about 25 to about 1000 nucleotides in length. In one embodiment, the donor fragments are single-stranded. Donor fragments can be such that they are derived from a single

strand of a parent polynucleotide. In one embodiment, the single-stranded regions are filled in using a polymerase. In one embodiment, the hybridized fragments are ligated using *Taq* DNA ligase or T4 DNA ligase. In a particular embodiment, the steps of hybridizing, filling in and ligating are repeated, such that one or more chimeric polynucleotides is used to generate the populations of scaffold or donor fragments. In one aspect, at least one of the fragments of the scaffold or donor populations contains at least one region of random sequence.

In another embodiment, the invention is directed to a method for preparing a population of scaffold fragments, including the steps of: amplifying an oligonucleotide of interest in a polymerase chain reaction, such that the 5' terminus of a first primer contains a 5' phosphate and the 5' terminus of a second primer is devoid of a 5' phosphate; contacting the amplified oligonucleotide with lambda exonuclease under conditions wherein oligonucleotides having a 5' phosphate are digested, leaving single-stranded oligonucleotides; and fragmenting the single-stranded oligonucleotides, thereby preparing a population of scaffold fragments.

In another embodiment, the invention is directed to a method for forming a chimeric polynucleotide including the steps of: treating a library of oligonucleotide fragments derived from a parent polynucleotide of interest and allelic variations thereof, wherein the population of fragments comprises a first population of oligonucleotides derived from one strand of the parent polynucleotide and allelic variations thereof and oligonucleotides of a second population wherein oligonucleotides are synthesized *in vitro* and derived from the other strand of the known parent polynucleotide and allelic variations thereof under conditions such that oligonucleotides of the first population can hybridize to oligonucleotides of the second population to form a gapped homoduplex; treating the gapped homoduplex with a polymerase, wherein polynucleotide strand extension produces a double-stranded polynucleotide comprising at least one nicked strand; and treating the nicked polynucleotide with a ligase, thus forming a full-length polynucleotide. In a particular embodiment, the invention is directed to a method of forming a single-stranded chimeric polynucleotide according, such that the oligonucleotides of the second population do not contain a 5' phosphate group, and includes the step of

removing the oligonucleotides of the second population after ligation. In a different embodiment, a single-stranded chimeric polynucleotide is formed using oligonucleotides of the second population that do not contain a 3' hydroxyl group. In the cases where a single-stranded chimeric polynucleotide is formed, scaffold fragments can be removed from the single-stranded chimeric polynucleotide after the ligation step. Single-stranded chimeric polynucleotides can be amplified in a nucleic acid amplification reaction to thereby produce more than one copy of a double-stranded chimeric polynucleotide. In one embodiment, at least one self-priming heteroduplex is a gapped heteroduplex including single-stranded sequences separated by double-stranded sequences. The gapped homoduplex can be full length. In one embodiment, the known parent sequence is from about 1 kilobase to about 5 kilobases in length. In another embodiment, the known parent sequence is from about 2 kilobases to about 25 kilobases in length. One aspect of the invention includes an additional recombination step between the chimeric polynucleotide and a parent molecule or allelic variation thereof.

In one embodiment, the invention is directed to a library of chimeric polynucleotides comprising more than one chimeric polynucleotides formed according to the methods described herein. The oligonucleotides of the second population can be derived from regions of sequence identity between parent polynucleotides and allelic variations thereof. In one embodiment, the gapped homoduplex can contain polymorphic sites in at least one double-stranded region of the homoduplex. In another embodiment, the gapped homoduplex can contain at least one polymorphic site in the gapped region of the gapped homoduplex.

In another embodiment, the invention is directed a method for directed evolution including the steps of: forming a library of chimeric polynucleotides by: contacting a first population of oligonucleotides with a second population of oligonucleotides, wherein the sequences of the first and second oligonucleotide populations are complementary to one another, under conditions such that oligonucleotides of the first population can hybridize to oligonucleotides of the second population to form a gapped homoduplex; treating the gapped homoduplex with a polymerase, such that polynucleotide strand extension produces a nicked

polynucleotide; treating the nicked polynucleotide with a ligase, such that nicks are ligated; and screening the library of chimeric polynucleotides for a characteristic of interest. In one embodiment, the oligonucleotides of the first population and the oligonucleotides of the second population are derived from a known polynucleotide of interest. In one aspect, the steps are repeated using the chimeric polynucleotide as the known polynucleotide of interest in the subsequent round of directed evolution. In a particular embodiment, the steps are repeated from about 2 to 50 times using a screened population of chimeric polynucleotides as the parent polynucleotides used to generate scaffold and donor fragments in a subsequent round of directed evolution. In one embodiment, the oligonucleotides of the second population do not contain 5' phosphate groups. In another embodiment, the oligonucleotides of the second population do not contain 3' hydroxyl groups. In a particular embodiment, the screening step includes screening the function of the transcribed and/or translated products of the library of chimeric polynucleotides. One aspect of the invention involves cloning the library of chimeric polynucleotides into a suitable vector prior to the screening step.

In a particular embodiment, the methods for directed evolution described herein include: cloning the chimeric polynucleotides into expression vectors; transforming a suitable cell line with the cloned chimeric polynucleotides; inducing expression of the cloned chimeric polynucleotide; assaying the expressed product for a characteristic of interest; and selecting the chimeric polynucleotide that expressed products with an improved characteristic of interest. In another embodiment, the methods for directed evolution described herein include: transcribing and translating the chimeric polynucleotide *in vitro*; assaying the transcribed and translated products for a characteristic of interest; and selecting the chimeric polynucleotide that lead to transcribed and translated products with an improved characteristic of interest.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of an embodiment of the invention, as illustrated in the accompanying drawings.

Figure 1 is a schematic diagram of one embodiment of the present invention.

Figures 2A and 2B depict synthetic genes and oligonucleotides. 2A. Alignment of TCI and PCI sequences altered to represent common *E. coli* codon usage. SNPs, DiPs and TriPs are as indicated. 2B. Oligonucleotides used for *in vitro* recombination of TCI and PCI genes. Restriction sites for cloning are underlined. Degenerate positions are shown in alternative bases in parentheses. Permutations of DiPs and TriPs are underlined. Note that the primers anneal to the genes in panel A and represent only the top strand.

Figures 3A and 3B are alignments of the human and mouse EGF coding sequences. 3A. Alignment of the unmodified human and mouse EGF coding sequences (with flanking engineered sequences and restriction cleavage sites). Identical positions are marked by dashes in the mouse sequence. Amino acid residue polymorphisms are indicated below the alignment. 3B. Alignment of the genes after design modifications to minimize genetic differences without altering the information content of the encoded polypeptides.

Figures 4A-C are schematic diagrams depicting different shuffling strategies. 4A. PARSed DNA shuffling of the mouse and human EGF polymorphisms. 4B. PARSed DNA shuffling of EGF polymorphisms from five species (human, mouse, rat, horse and pig). 4C. Heteroduplex DNA shuffling of degenerate oligonucleotides.

Figures 5A and 5B are schematic diagrams representing PARSed DNA shuffling products. 5A. Two-gene PARSed shuffling. Black blocks contain only human nucleotide polymorphisms. White blocks contain only mouse polymorphisms. 5B. Five-gene PARSed shuffling. Blocks containing codons from each of the five mammalian species are uniquely shaded. Regions containing polymorphisms that cannot be assigned to a single parent are left unshaded. Unambiguous crossovers in such regions are indicated by vertical lines.

Figure 6 is a graph showing the frequency of reassortment of polymorphisms. DNA sequence information for 8 unselected clones indicated representation of each allele at every polymorphic position. The number of human vs. mouse alleles at any one position ranged from 2 to 6 (fraction = 0.25 to 0.75) and clustered near the theoretically ideal value of 0.5.

DETAILED DESCRIPTION OF THE INVENTION

The methods of the present invention facilitate the generation of chimeric polynucleotides and do not require hybridizing donor fragments to a target- or full-length template. Rather, a first population of hybridizing donor fragments, *e.g.*, oligonucleotides, are assembled using a second population of scaffold fragments, *e.g.*, oligonucleotides, to form a double-stranded chimeric polynucleotide in which one or both strands are chimeric. One strand of the chimeric double-stranded polynucleotide can comprise, for example, scaffold fragments and regions between the scaffold fragments that were filled-in during the process; the opposite strand can also comprise donor fragments and regions between the donor fragments that were filled-in during the process. Because the chimeragenesis process of the present invention does not rely upon a contiguous, full-length template, it is unnecessary to modify a template to facilitate its removal.

"Chimeric polynucleotides", as used herein, contain nucleotide sequences from multiple related sequences or otherwise similar polynucleotides, referred to herein as "parent polynucleotides." "Full-length," as used herein to describe polynucleotides, is a relative term meaning the product is about the same length as the parent polynucleotide. In one embodiment, the scaffold is made up or otherwise designed, generated or derived of fragments from one strand, *e.g.*, the top strand, of a parent polynucleotide, *e.g.*, a template polynucleotide. In another embodiment, the scaffold is formed without reference to a particular strand of a parent polynucleotide, but the scaffold fragments are nonetheless complementary to the donor fragments.

The methods described herein comprise process steps involved in the formation of chimeric polynucleotides. Reference is now made to Figure 1 which depicts schematically the steps utilized by one embodiment of the present invention

in forming the double-stranded chimeric polynucleotide, wherein a population of scaffold fragments are used to assemble the hybridizing population of donor fragments. A polynucleotide of interest, *e.g.*, a gene, 10 is used to prepare a population of single-stranded scaffold fragments 20. A population of donor fragments 30 is assembled into hybridization complexes 40 with the scaffold fragments. In some cases, overlaps occur between donor fragments and/or scaffold fragments, thus creating "flaps" 50. The term "flaps" is intended to include the unhybridized terminal portions of a fragment that is otherwise hybridized to another fragment. Overlaps can occur between hybridized scaffold fragments or hybridized donor fragments. In other cases, regions between the hybridized fragments remain single-stranded, thus creating "gaps" between the fragments 60. Flaps can be trimmed and gaps can be filled-in prior to the generation of a contiguous chimeric polynucleotide 70. A contiguous double-stranded chimeric polynucleotide can be generated by ligating the assembled oligonucleotides 80. The method of the present invention can further include repeating the method using at least one chimeric polynucleotide or fragment thereof as the scaffold fragments or donor fragments.

Figure 4 is a schematic diagram depicting different shuffling strategies. Figure 4A depicts PARtially Scaffolded (PARSed) DNA shuffling of the mouse and human EGF polymorphisms. Three degenerate 5'-phosphorylated top strand (TS1-3) and two non-phosphorylated partial scaffold (PS1/2) oligonucleotides were synthesized to contain all the amino acid polymorphisms of the parental mouse and human *EGF* genes and silent modifications. Arrows indicate the position and number of alternative codons. Arrows opposite gaps in the top strand indicate SNPs that were incorporated by primer extension, using the scaffold oligonucleotides as templates. Boxes indicate where alternative codons were synthesized in separate reactions in order to minimize degeneracy. Dotted vertical lines indicate homoduplex base pairing between top strand degenerate positions and their complementary bases in the scaffolds. Bold numbers above or below each oligonucleotide indicate its length in nucleotides. Numbers between oligonucleotides indicate the number of nucleotides available for hybridization up to the first degenerate position. Spaces between top strand oligonucleotides represent

nicks, except where underlined numbers indicate gap lengths. Figure 4B shows PARSed DNA shuffling of EGF polymorphisms from five species (human, mouse, rat, horse and pig). The two gray arrows indicate inclusion of amino acids not present in the parental genes. Figure 4C shows a method for heteroduplex DNA shuffling of degenerate oligonucleotides. Oligonucleotides with polymorphisms from two parental genes were annealed to a full length template representing either of the two parental genes. Arrows topped with an "x" indicate the heteroduplex mismatches closest to the oligonucleotide ends. As used herein, "homoduplex" polynucleotides refer to hybridized strands that contain only Watson-Crick base pairs, *i.e.*, they do not contain "mismatches." "Heteroduplexes" are hybridized polynucleotide molecules that contain at least one mismatch.

The chimeric polynucleotides described schematically in Figures 1 and 4 include crossovers. As used herein, "crossover" refers to an event that leads to strand switching. As used herein, "strand switching" describes a nucleotide sequence such that the sequence is identical to a reference polynucleotide up to the "switch" or "crossover" site, and the sequence downstream of the crossover site is identical to a different reference polynucleotide. Strand switching is a description of nucleotide sequence and is not necessarily indicative of a physical switching of strands; see, for example, Figures 3 and 5 which depict chimeric sequences containing crossovers.

In embodiments, the population of scaffold fragments is derived from one or more allelic versions of the gene of interest. As used herein, "allelic version" refers to a polynucleotide with a sequence similar to a reference polynucleotide, *e.g.*, a wild-type gene. The allelic version typically has a different sequence at one or more "polymorphic sites" with respect to the reference polynucleotide. As used herein, "polymorphic sites" refer to those positions in a sequence where, in a population of related polynucleotides, more than one sequence occurs. As used herein, "degeneracy" refers to either the number of allelic variants at a polymorphic site, or the number of polymorphic sites contained in a nucleotide sequence; a higher level of degeneracy corresponds to a greater number of variants possible at a polymorphic site or a greater number of polymorphic sites in a particular sequence. In contrast,

among the population of polynucleotides, the sequences are identical at non-polymorphic sites. Polymorphic sites can exhibit single nucleotide polymorphisms (SNPs), dinucleotide polymorphisms (DiPs) and trinucleotide polymorphisms (TriPs) or combinations thereof. Expected frequencies of crossovers can be calculated based on the fold degeneracy at a particular polymorphic site. For example, if the theoretical limit of crossover events is reached, one would expect to have an equal chance (determined by the input of the donor and scaffold fragments) of having a particular allelic variant at a particular polymorphic site irrespective of the allelic variant present at a different polymorphic site. For example, if a two-fold degenerate polymorphic site is included in a shuffling reaction such that an equal number of each variant is used to generate scaffold and donor fragments, it would be expected that 50% of the resultant chimeric polynucleotides would contain each allelic variant. This 50% value is independent, *i.e.*, does not display "genetic linkage," of the specific allelic variant at a different polymorphic site.

A population of scaffold fragments can be contacted with a population of donor fragments. The interactions between donor fragments and scaffold fragments occur by the process of hybridization. Thus, some degree of complementarity between scaffold fragments and donor fragments must exist to allow for such interactions. Further, since the interactions between scaffold and donor fragments are dependent on base-pairing, in a particular embodiment, scaffold fragments are derived from a polynucleotide strand complementary to the strand from which the donor fragments are derived. In one embodiment, scaffold fragments are derived from a reference polynucleotide or "template" polynucleotide. In another embodiment, scaffold fragments are derived from the strand of the allelic version that is complementary to the strand that is used to generate the donor fragments.

The scaffold fragments typically comprise single-stranded molecules, having minimal complementarity with each other. As used herein, "minimal complementarity" between scaffold fragments means that a scaffold fragment will hybridize with other fragments, *e.g.*, donor fragments, to form a duplex with a higher melting temperature than a duplex formed by the scaffold fragment hybridizing to another scaffold fragment. In a preferred embodiment, the single-stranded scaffold

fragments are prepared, synthesized, designed, generated or otherwise derived from one strand, *e.g.*, the top strand of a polynucleotide of interest. Single-stranded scaffold fragments can be prepared by nicking a single-stranded polynucleotide of interest or by denaturing a nicked double-stranded polynucleotide. Single-stranded polynucleotides can be made by denaturing double-stranded polynucleotides. "Denaturing," as used herein, refers to the process of physically separating the single strands of nucleic acid by disrupting base-pairing interactions between complementary strands. Such methods of denaturing double-stranded polynucleotides are well known in the art.

Scaffold fragments can be made, for example, using enzymatic techniques, physical techniques or chemical synthesis techniques. For example, single-stranded polynucleotides can be synthesized by PCR amplification of a polynucleotide of interest using primers wherein one primer contains a terminal 5' phosphate and the other primer does not contain a terminal 5' phosphate. The amplified product can then be treated such that nucleic acid having a terminal 5' phosphate or molecules devoid of a terminal 5' phosphate are preferentially destroyed. In one embodiment, the amplified product is treated with lambda exonuclease to degrade the phosphorylated strand. Single-stranded polynucleotides can also be made by inserting a polynucleotide of interest into M13 phage and performing first strand synthesis using methods well known in the art. The single-stranded polynucleotides can be fragmented by enzymatic, chemical, or physical techniques known in the art in order to generate single-stranded scaffold fragments.

The scaffold fragments can be generated from a larger polynucleotide and treated to form fragments or single-stranded fragments. In embodiments, double-stranded polynucleotides are fragmented such that fragments of one strand form the population of donor fragments and fragments of the complementary strand form the population of scaffold fragments. In this embodiment, the double-stranded polynucleotides that are fragmented are, preferably, different allelic versions of each other.

The scaffold fragments can be of any length which is less than that of a full-length polynucleotide of interest, *e.g.*, less than the full-length of the corresponding

wild-type gene. Preferably, scaffold fragment are considerably shorter than the full-length polynucleotide, most preferably not more than 30% of its length. In one embodiment, the scaffold fragments can be from about 20 to about 1500 nucleotides in length. In another embodiment, the scaffold fragments can be from about 25 to about 1000 nucleotides in length or from about 100 to about 1000 nucleotides in length. The scaffold fragments can be at least about 40 nucleotides in length, at least about 100 nucleotides in length, or at least about 1000 nucleotides in length. The scaffold fragments can be less than about 25 percent of the desired length of the chimeric polynucleotide products, or about 15 or 20 percent or less of the desired length of the chimeric polynucleotide products. Without wishing to be bound by theory, while the use of longer scaffold fragments can facilitate the formation of target length chimera in the absence of thermocycling or multiple rounds of annealing and denaturing, shorter scaffold fragments can facilitate the number of crossovers. As used herein, "target-length chimera" refers to the approximate length of a hypothetical chimeric polynucleotide having the desired properties. Target length can be estimated based on the length of the polynucleotide of interest or reference polynucleotide as described herein.

The scaffold fragments allow for the assembly of donor fragments into an ordered duplex with at least one scaffold fragment. Typically, the scaffold fragments are selected such that they are related to the parent polynucleotides of interest, *e.g.*, genes, which are allelic versions of each other, that are used to generate the donor fragments. In another embodiment, the scaffold fragments are derived from a reference polynucleotide of interest (a "template") and the donor fragments are derived from allelic versions of the template or a combination of allelic versions and the template. In a particular embodiment, the scaffold fragments are derived from a particular strand of a duplex polynucleotide. For example, the scaffold fragments can be derived from the sense or top strand, and the donor fragments can be derived from the antisense or bottom strand. The polynucleotide of interest can comprise a gene, either a genomic copy or cDNA (or intronless) copy. The polynucleotide of interest can comprise more than one coding sequence. For

example, the polynucleotide of interest can comprise an operon including regulatory regions, either as a single contiguous molecule or as more than one molecule.

The nucleic acids for use as either scaffold or donor fragments can be synthetically manufactured or isolated from any suitable source of nucleic acid. The scaffold or donor fragments of the present invention can comprise DeoxyriboNucleic Acid (hereinafter "DNA"), or RiboNucleic Acid (hereinafter "RNA") DNA or RNA can comprise natural bases, *e.g.*, adenine, thymine, cytosine, guanine or uracil; analog bases, *e.g.*, inosine, bromouracil or nitroindole; chemically altered bases, *e.g.*, biotin labeled or digoxygenin labeled bases; or a combination thereof provided that the resulting double-stranded chimeric polynucleotide can be replicated. Scaffold fragments can be such as the cannot be ligated, *e.g.*, they lack 5' phosphate groups, or they can be such that they can not be extended, *e.g.*, they lack 3' hydroxyl groups.

Further, polynucleotides used to generate the scaffold or donor fragments, or the fragments themselves can be isolated from an organism, such as, for example, a eubacterial, archeal, eukaryotic or viral organism. These organisms can be amplified, enriched or isolated and grown in culture, or can be used directly from environmental sources. Environmental sources include soil samples, water samples from fresh water sources or salt water sources, polluted sites, waste treatment sites and sources including extreme condition sources such as permafrost sources, high altitude sources, high pressure sources and geothermal sources such as volcanic sources, hot springs and hydrothermal vent sources. Sources of nucleic acid also include tissue or bodily fluid samples from an organism, such as a human samples and include human genomic DNA. The nucleic acid of a tissue or bodily fluid sample can include nucleic acid of the organism, such as chromosomal, episomal or transcribed nucleic acid, or can be nucleic acid of the flora, such as fungal, bacterial, viral or parasitic organisms present in the sample. The sample can further be fresh, fossil or archival.

It is understood that the nucleic acid isolated from these sources can be produced in the form of a genomic or cDNA library using methods well known in the art. In the case of cDNA, RNA or preferably polyA$^+$ RNA or mRNA is isolated from a sample, and converted into double-stranded DNA (cDNA) according to

standard methods, well known in the art. In one embodiment, a cDNA library is prepared from a sample of interest that expresses the desired phenotype. In another embodiment, the cDNA library can be enriched for sequences of interest prior to use as oligonucleotides. The cDNA library can be subjected to subtractive hybridization against a suitable sample of nucleic acid using subtractive hybridization techniques well known in the art. A suitable sample of nucleic acid includes, for example, nucleic acid from a reference strain of bacteria. In one embodiment, sequences that are common between the cDNA library and the sample nucleic acid are allowed to hybridize to each other and double-stranded nucleic acids are then removed from the pool. In this way, sequence present in multiple copies and sequences that are common between the two populations are removed, effectively enriching for low abundance or unique sequences. For example, a library of donor fragments prepared according to the method as described in "Generating Single-Stranded Oligonucleotide Libraries with Minimal Complementarity and Uses Therefore" by Joseph J. Arensdorf and Wayne M. Coco, United States Application No. 09/691,873 filed October 19, 2000.

The scaffold or donor fragments of the present invention can be isolated from any suitable source of oligonucleotides as described herein. Methods of choosing and/or isolating nucleic acids from suitable sources of nucleic acid are well known in the art. In another embodiment of the present invention, the scaffold or donor fragments (or both) can be produced *in vitro* using enzymatic or chemical means. Methods of *in vitro* production of nucleic acid sequence are well known in the art.

The scaffold or donor fragments can include one or more regions with functional characteristics or structural motifs of the parent polynucleotides. The scaffold or donor fragments can comprise all or a portion of a region with functional characteristics or structural motifs. These regions can include nucleic acid structural motifs, protein binding domains, metal binding domains, nucleic acid binding domains, domains with enzymatic activity, or fragments of these domains. These regions can include ribozymes, deoxyribozymes, promoters, enhancers, origins of replication, open reading frames, or fragments thereof. These regions can encode

aptamers, wherein aptamers are small single- or double-stranded DNA or RNA molecules that bind specific molecular targets (Bock *et al.*, *Nature 355*:564-566, 1992; Ellington and Szostak, *Nature 346*:818-822, 1990; Werstuck and Green, *Science 282*:296-298, 1998).

The scaffold or donor fragments of the present invention can also include regions of sequence that are not known to have any particular function. These regions can be selected from any known source of nucleic acid sequence, including sequences synthesized *in vitro*, or these regions can be of random or partially random sequence. Partially random sequences can be generated by synthesizing a oligonucleotide based on a known sequence, except that a portion of the sequence is randomized (*e.g.*, randomizing the last 50 nucleotides), or wherein certain positions within the sequence are randomized (*e.g.*, randomizing particular codon(s) of a coding sequence) or wherein certain bases are randomized (*e.g.*, randomizing all adenines). These regions can further encode proteins or domains of proteins including folding structures or structural motifs; binding domains such as protein binding domains, metal binding domains, co-factor binding domains, lipid binding domains and nucleic acid binding domains; domains with enzymatic function; sites for allosteric or competitive inhibition and the like; or fragments of these domains. These regions can also include amino acid sequences that are not known to have any particular function or can be randomized amino acid sequence.

The parent polynucleotides can be fragmented while in double-stranded or single-stranded form. Preferred methods for cleaving, *e.g.*, fragmenting parent polynucleotides in order to generate populations of donor and scaffold fragments are those methods which produce fragments without particular sequence patterns. In one embodiment of the present invention, a population of fragments is created by randomly fragmenting parent polynucleotides.

The parent polynucleotides, scaffold fragments or donor fragments are generated using chemical, physical or enzymatic techniques. Chemical techniques of fragmenting polynucleotides can include techniques that utilize pH extremes, hydroxyl radical formation, chemical radical formation, chemical catalysis or a combination thereof. Methods of fragmenting polynucleotides by chemical

techniques can be used to generate defined or undefined ends. Techniques are well known in the art such that polynucleotides can be hydrolyzed after defined bases (*e.g.*, only after guanines), or hydrolyzed to generate undefined termini. For example, exposure of polynucleotides to extreme pH (*e.g.*, acidic pH or basic pH) can generate fragments with undefined termini. Additionally, hydroxyl radicals (*e.g.*, generated using Fenton or Udenfriend reagent) react with the deoxyribose in DNA, resulting in cleavage of the DNA strand. The result is near uniform cleavage at any base within a target polynucleotide, and the frequency of cleavage can be regulated. In addition to fragmenting polynucleotides by chemical techniques, physical techniques, such as heating, freezing, using ionizing radiation and shearing can be employed.

Yet another approach to creating a population of fragments involves the use of enzymatic techniques. These methods can include the use of any suitable enzyme such as a nucleic acid polymerizing enzyme or a nuclease. For example, a polymerase can be used to synthesize oligonucleotides of variable length. Where fragments are generated by parent polynucleotide-dependent synthesis, conditions of synthesis can be chosen such that the polymerase arbitrarily falls off the polynucleotide or otherwise terminates synthesis at arbitrary points along the polynucleotide. This approach allows for oligonucleotides to be generated with arbitrary sequence alterations (*e.g.*, "error-prone" methods). Another method for using polymerases to generate a fragmented population of oligonucleotides uses polymerases that are known to have exonuclease activity under conditions permitting exonuclease activity. Such enzymes include, for example, T4 DNA polymerase, PolI, PolIII, *Pfu* polymerase and Klenow polymerase.

Still another method for enzymatically generating a population of oligonucleotides with undefined termini involves removing bases or generating adducts in an oligonucleotide using techniques well known in the art. For example, specific bases in oligonucleotides can be removed or adducted by many well known chemical methods to result in either abasic sites or chemically altered bases. These sites can be produced, for example, between 15 and 5000 bases apart (Kunkel *et al.*, *Meth. Enzymol. 154*:367-382, 1987). Strand cleavage of the phosphodiester bond at

those modified sites can then be effected using chemicals such as piperidine, or enzymes such as abasic lyases or abasic endonucleases.

Another enzymatic method for creating a fragmented population of oligonucleotides uses endonucleases having sequence-specific recognition sites. Such enzymes are known as "restriction endonucleases" and are commercially available. A fragmented population of oligonucleotides can be generated by performing a limited or incomplete digestion of the parent polynucleotides. Additionally, oligonucleotides having undefined termini can be generated by using non-specific endonucleases such as mung bean exonuclease, S1 nuclease or DNase I. In another embodiment of generating oligonucleotides having undefined ends, exonucleases such as ExoIII or ExoVII can be used to non-specifically trim oligonucleotide sequences.

The fragmented population of oligonucleotides can include oligonucleotides containing random or partially random sequence. The population of fragments can include molecules generated using any one of the above described methods or combinations thereof. The term "random" as used herein is intended to reflect an absence of preselection. Such absence can be of any degree; it need not be a total absence of preselection, nor does the term indicate a requirement for an absence of preference or bias. The term can be used to describe populations of oligonucleotides, sequences, events, processes, states or conditions, or other such terms. Such compositions can range over a span of values and any one component can occupy any of these values. For example, a population of oligonucleotides that is generated by the digestion of two polynucleotides with a restriction enzyme is a "random population" when the particular oligonucleotides formed by the process are not preselected, for example, during a partial digestion. This is true even when the gene sequences are known and the restriction enzyme preferentially cleaves a particular site. Sequences can be random if at least one position in the sequence is not specifically defined (for example, if at least one position of an oligonucleotide could be and is either one of two or more nucleotides, *e.g.*, a polymorphic site). The randomly fragmented population of oligonucleotides can include oligonucleotides

wherein a portion of the oligonucleotides comprise random or partially random sequence as described herein.

In a preferred embodiment, the scaffold population is not randomly produced but is designed to optimize crossover events. Such a scaffold population can provide either chimeragenesis or a lack of chimeragenesis, but will correspond to the whole or a substantial, although not necessarily contiguous, length of the polynucleotide of interest, *e.g.*, gene of interest. For example, scaffold fragments can be synthetically produced to each contain complementary sequences, *i.e.*, termini, to two donor fragments. Thus, each terminus of the scaffold fragment can hybridize to a different donor fragment and each donor fragment terminus (with the exception of flaps) can hybridize to a different scaffold fragment terminus. The donor fragments can be randomly generated or specifically designed to introduce chimeragenesis. The scaffold can be designed to have identity to the termini of the donor fragments to provide the desired cross-overs and gaps to correspond to the desired mutations or chimeragenesis.

The population of scaffold fragments of the present invention includes oligonucleotides that are typically shorter than target length chimera. The target length *e.g.*, length of resulting double-stranded chimera, can be from about 50 to about 100,000 nucleotides in length. In particular embodiments, the target length can be from about 100 to about 50,000 nucleotides in length; from about 200 to about 10,000 nucleotides in length; from about 500 to about 5,000 nucleotides in length or from about 1,000 to about 3,000 nucleotides in length.

The population of donor fragments includes oligonucleotides from about 5 to about 50,000 nucleotides length. In more particular embodiments, the population of donor fragments includes oligonucleotides from about 10 to about 10,000 nucleotides in length, from about 15 to about 5,000 nucleotides in length, from about 20 to about 2,500 nucleotides in length, from about 25 to about 1,000 nucleotides in length, or from about 40 to about 200 nucleotides in length. The donor fragments can be at least about 40 nucleotides in length, at least about 100 nucleotides in length or at least about 1000 nucleotides in length.

The scaffold fragments guide the hybridizing donor fragments and form a double-stranded chimeric polynucleotide. Where the donor fragments are double-stranded molecules, they can be denatured prior to hybridization with the scaffold fragments. Methods of denaturing and annealing donor fragments sequences, are well known in the art. In a particular embodiment, the donor fragments are single-stranded and from the opposite strand compared to the scaffold fragments. "Opposite strand" refers to, for example, the donor fragments being derived from the antisense or bottom strand of a duplex polynucleotide when the scaffold fragments are derived from the top or sense strand. In a particularly preferred embodiment, the scaffold fragments are derived from the top strand, and the donor fragments are derived from the bottom strand. In another embodiment, the population of donor fragments or the scaffold fragments, or both share minimal complementarity with members of the same population. Minimal complementarity can be achieved by treating the members of the population such that they do not hybridize to each other. Alternatively, minimal complementarity can be achieved by selecting the members of the population such that they do not hybridize to each other, *e.g.*, forming the population by cleaving one strand of a polynucleotide of interest. It is clear that one of skill in the art can skew the availability of a given donor fragments to hybridize a scaffold fragment by including an oligonucleotide capable of hybridizing to said donor fragment. In one embodiment, a region of the scaffold is hybridized to complimentary sequences by providing oligonucleotides complimentary to a specific scaffold sequence. In this manner, a region of the scaffold can be specifically retained in the resultant double-stranded chimeric molecule. Conversely, defined oligonucleotides can be added in greater quantities to the population of donor fragments in order to preferentially hybridize the defined oligonucleotides to the scaffold at particular regions or positions in order to introduce desired mutations or in order to protect sequences on the scaffold from changes that might be introduced by the arbitrarily fragmented population of donor fragments.

The donor fragments can also be single-stranded, and in a particular embodiment, are derived from the opposite, *e.g.*, complementary, strand to the strand of a duplex polynucleotide from which the scaffold fragments are derived. As

used herein, "derived" refers to a sequence identical to a sequence contained in a reference polynucleotide except at polymorphic sites. Thus, when the donor and scaffold fragments contact each other, a hybridized complex can form, which generally comprises at least one donor fragment hybridized to at least one scaffold fragment. In a particular embodiment, the hybridized complex comprises at least two donor fragments hybridized to at least one scaffold fragment. Single-stranded regions remaining between adjacently hybridized fragments, herein referred to as "gaps," can be filled in, *e.g.*, using a polymerase. Where there is a 3' overhang, the overhang can be filled in by adding a primer that hybridizes at or near the free terminus of the 3' overhang such that polymerization during gap filling can proceed on the 3' overhang. Adjacently hybridized fragments can then be ligated to form a double-stranded polynucleotide comprising chimeric polynucleotides.

The present invention allows donor fragments of interest and scaffold fragments to be incorporated into a larger molecule to form one or more double-stranded chimeric polynucleotides. In one embodiment, polynucleotides that are not otherwise easily manipulated (*e.g.*, large polynucleotide chains), can be separately manipulated as oligonucleotides and rejoined by contacting the oligonucleotides with single stranded scaffold fragments to form a hybridized complex. For example, random mutagenesis using PCR is most effective on smaller DNA fragments, such as 1 kilobase or less in length. A large polynucleotide can be cleaved into fragments of about one kilobase, randomly mutagenized using PCR, and then denatured. Denatured and mutagenized fragments can be contacted with scaffold fragments to form a hybridized complex, filled in and ligated as described herein. The template scaffold can be derived from the original oligonucleotide, or can be modified as described herein. For example, the scaffold fragments can be mutagenized or can have added or deleted regions or domains as compared to the starting polynucleotide.

It is clear to one of skill in the art that the method of the present invention can be carried out under a range of reaction conditions and hybridization conditions. Conditions can be selected based on the amount of similarity or differences between the oligonucleotides and the template. In one embodiment of the present invention,

the donor fragments are hybridized or annealed to the scaffold fragments under conditions of low stringency.

A general description of stringency for hybridization and wash conditions is provided by Ausubel, F.M. *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Assoc. and Wiley-Interscience 1987, & Supp. 49, 2000, the teachings of which are incorporated herein by reference. Factors such as probe length, base composition, percent mismatch between the hybridizing sequences, temperature and ionic strength of reactions, hybridizations and washes influence the stability of nucleic acid hybrids. Thus, stringency conditions sufficient to allow hybridization of donor and scaffold fragments to form hybridization complexes can vary significantly and still allow for the generation of at least one chimeric polynucleotide. The energetics favoring hybridization indicate that longer stretches of homology are more favorable. Thus, when either short sequences are involved or there is limited potential for standard Watson-Crick base-pairing, hybridization conditions can be adjusted to a lower stringency to allow for hybridization. Typically, adjusting hybridization and wash conditions is done by, for example, adjusting the ionic strength of the reaction mixture or adjusting the temperature at which the hybridization is performed. In addition, certain purified proteins, such as the *E. coli* RecA protein, aid in homologous base pairing and can be included to facilitate hybridization of polynucleotide strands.

While not wishing to be bound by theory, typically, when two fragments anneal to form a hybridization complex, one or two single-stranded termini remain. These single-stranded termini can anneal to additional fragments from the mixture by altering hybridization conditions to favor the annealing of multiple fragments in a hybridization complex. To facilitate the hybridization of fragments having low homology, the donor and scaffold fragments can be allowed to anneal (hybridize) at 50°C. In another embodiment, the donor and scaffold fragments can be allowed to anneal at 60°C or at 70°C. To facilitate the hybridization of multiple donor fragments and scaffold fragments in a hybridization complex, the donor and scaffold fragment mixture can be held at the annealing temperature for at least about 30 seconds. In another embodiment, the donor and scaffold fragment mixture can be

held at the annealing temperature for at least about 1 minute, 2 minutes, 5 minutes, 15 minutes, 30 minutes, 1 hour, 5 hours, 10 hours or 24 hours. Combinations of annealing temperature and incubation time at the annealing temperature can be used to facilitate the formation of hybridization complexes comprising multiple donor and scaffold fragments.

Alternatively, conditions for stringency are as described in WO 98/40404, the teachings of which are incorporated herein by reference. In particular, examples of "highly stringent," "stringent," "reduced," and "least stringent" conditions are provided in WO 98/40404 in the Table on page 36. Examples of stringency conditions are shown in the table below which is from WO 98/40404. Highly stringent conditions are those that are at least as stringent as, for example, conditions A-F; stringent conditions are at least as stringent as, for example, conditions G-L; and reduced stringency conditions are at least as stringent as, for example, conditions M-R.

Stringency Condition	Oligonucleotide Hybrid	Hybrid Length (bp)‡	Hybridization Temperature and Buffer†	Wash Temperature and Buffer†
A	DNA:DNA	≥ 50	65°C; 1xSSC -or- 42°C; 1xSSC, 50% formamide	65°C; 0.3xSSC
B	DNA:DNA	<50	T_B*; 1xSSC	T_B*; 1xSSC
C	DNA:RNA	≥ 50	67°C; 1xSSC -or- 45°C; 1xSSC, 50% formamide	67°C; 0.3xSSC
D	DNA:RNA	<50	T_D*; 1xSSC	T_D*; 1xSSC
E	RNA:RNA	≥ 50	70°C; 1xSSC -or- 50°C; 1xSSC, 50% formamide	70°C; 0.3xSSC
F	RNA:RNA	<50	T_F*; 1xSSC	T_F*; 1xSSC
G	DNA:DNA	≥ 50	65°C; 4xSSC -or- 42°C; 4xSSC, 50% formamide	65°C; 1xSSC
H	DNA:DNA	<50	T_H*; 4xSSC	T_H*; 4xSSC
I	DNA:RNA	≥ 50	67°C; 4xSSC -or- 45°C; 4xSSC, 50% formamide	67°C; 1xSSC
J	DNA:RNA	<50	T_J*; 4xSSC	T_J*; 4xSSC
K	RNA:RNA	≥ 50	70°C; 4xSSC -or- 50°C; 4xSSC, 50% formamide	67°C; 1xSSC
L	RNA:RNA	<50	T_L*; 2xSSC	T_L*; 2xSSC
M	DNA:DNA	≥ 50	50°C; 4xSSC -or- 40°C; 6xSSC, 50% formamide	50°C; 2xSSC
N	DNA:DNA	<50	T_N*; 6xSSC	T_N*; 6xSSC
O	DNA:RNA	≥ 50	55°C; 4xSSC -or- 42°C; 6xSSC, 50% formamide	55°C; 2xSSC
P	DNA:RNA	<50	T_P*; 6xSSC	T_P*; 6xSSC
Q	RNA:RNA	≥ 50	60°C; 4xSSC -or- 45°C; 6xSSC, 50% formamide	60°C; 2xSSC
R	RNA:RNA	<50	T_R*; 4xSSC	T_R*; 4xSSC

‡: The hybrid length is that anticipated for the hybridized region(s) of the hybridizing oligonucleotides. When hybridizing a oligonucleotide to a target oligonucleotide of unknown sequence, the hybrid length is assumed to be that of the hybridizing oligonucleotide. When oligonucleotides of known sequence are hybridized, the hybrid length can be determined by aligning the sequences of the oligonucleotides and identifying the region or regions of optimal sequence complementarity.

†: SSPE (1xSSPE is 0.15M NaCl, 10mM NaH_2PO_4, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes after hybridization is complete.

*T_B - T_R: The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, T_m(°C) = 2(# of A + T bases) + 4(# of G + C bases). For hybrids between 18 and 49 base pairs in length, T_m(°C) = 81.5 + 16.6($\log_{10}$[Na^+]) + 0.41(%G+C) - (600/N), where N is the number of bases in the hybrid, and [Na^+] is the concentration of sodium ions in the hybridization buffer ([Na^+] for 1xSSC = 0.165 M).

It is clear to one of ordinary skill in the art that the contacting and hybridization steps can be optimized using any suitable method of optimization that is established in the art of hybridization. These include, but are not limited to, techniques that increase the efficiency of annealing or hybridization from complex mixtures of oligonucleotides (*e.g.*, PERT; *Nucleic Acids Research 23*:2339-2340, 1995) or hybridization in different formats (*e.g.*, using an immobilized template or using microtiter plates; *Analytical Biochemistry 227*:201-209, 1995).

Any parent polynucleotide with sufficient sequence similarity to the scaffold can be used to generate the donor fragments of the present invention. As defined herein, "sufficient sequence similarity" means that the sequence of the oligonucleotide need not reflect the exact sequence of the scaffold. Conditions are chosen to allow such sequences (and those having low similarity or similar sequences interrupted with dissimilar sequences) to hybridize the scaffold, such that double-stranded chimeric polynucleotides are formed. For example, non-complementary bases or insertions or deletions can be interspersed in sequences.

Upon contacting donor and scaffold fragments with each other, at least one hybridized complex is formed. Where flaps (unhybridized termini), gaps (single-stranded regions) and/or nicks occur in the hybridized complex, they can be trimmed, filled and ligated. In a particular embodiment, immediately adjacent oligonucleotides are ligated to each other. The term "adjacently hybridized" is used herein to describe the relative positions of two scaffold fragments hybridized to the same donor fragment, or two donor fragments hybridized to the same scaffold fragment, at positions such that only single-stranded sequence is contained between the two fragments. The term "immediately adjacently hybridized" is used herein to

describe adjacently hybridized scaffold or donor fragments that abut each other, *e.g.*, no intervening single-stranded sequence is contained between the two hybridized fragments.

Typically, a trimming, polymerization, ligation (TPL) step follows the contacting and hybridization of the population of donor fragments to the scaffold fragments. The TPL step includes trimming flaps, polymerization to fill in gaps between adjacently hybridized fragments, and ligation to join immediately adjacently hybridized fragments.

The utility of trimming flaps is realized because, in certain cases, the population of donor and scaffold fragments can hybridize such that at least one terminus of at least one of the hybridized fragments is unhybridized. The term "flaps" is used herein to describe the unhybridized terminus of an otherwise hybridized fragment. Internal sequences can also remain unhybridized, thus forming "loops" (loops are observed, for example, during denaturation/renaturation experiments with cDNA and genomic genes in which genomic introns loop out since there is no corresponding cDNA sequence to which to hybridize). The "trimming" of flaps, used herein to refer to a process of removing just the flaps, leaving the hybridized portion of the fragment intact, can be incorporated into the method of the present invention. Flaps can be trimmed enzymatically, *e.g.*, utilizing polymerases with single-stranded exonuclease activity or other single-stranded endonucleases or exonucleases, or chemically. The step of trimming flaps can be performed prior to or concurrently with the additional steps of polymerization and ligation.

Depending on specific hybridization capabilities, fragments can hybridize such that segments of the fragments remain unhybridized, *i.e.*, "gaps" are created. Such gaps could prevent the final formation of template-length chimeric polynucleotide, so a polymerization step is used to fill in the gaps. In a particular embodiment, the termini of the fragments are hybridized and at least one internal segment of a hybridized fragment is not hybridized.

Polymerization can be achieved either chemically or enzymatically. For example, gaps between adjacently hybridized fragments can be filled using a

suitable nucleic acid polymerizing enzyme, *e.g.*, a "polymerase". Suitable polymerases are commercially available. In one embodiment, gaps are filled in using prokaryotic, eukaryotic or viral polymerases. The polymerase can be thermostable or not thermostable. The polymerases can optionally have proof reading ability. Suitable polymerizing enzymes include T4 DNA polymerase, *Taq* DNA polymerase, *Pfu* DNA polymerase, Pol I, Klenow and Klenow 3'→5'^{exominus}. (New England BioLabs, Beverly, MA). Typically, polymerases require a "primer" oligonucleotide that is extended by the polymerase in a process known as "strand extension." The polymerase reads the bottom strand and extends the primed top strand. A primer can be, for example, a short oligonucleotide that hybridizes to the bottom strand.

Control of enzymatic polymerization can be achieved, for example, by affecting the polymerase, *e.g.*, using a polymerase with altered processivity, or by affecting the template which is used by the polymerase during polymerization. For the purposes of the present invention, the gaps can be filled with or without the introduction of "errors" in comparison to the hybridized fragments.

In the method of the present invention, gaps between adjacently hybridized fragments can be separated by about 1,000 to about 100,000 template nucleotides. In other embodiments, the adjacently hybridized fragments are separated by about 500 to about 10,000 nucleotides; less than 1,000 nucleotides; less than 250 nucleotides; less than 50 nucleotides; or are separated by less than 25 nucleotides.

In another embodiment, gaps are filled in *in vivo*, wherein complexes containing oligonucleotides hybridized fragments are inserted or transformed into a suitable host cell. Gapped duplexes are examples of "self-priming" substrates for polymerases in the instances where the top strand contains an extendable 3' end and the single-stranded gap is used as the bottom strand that is read by the polymerase to extend the self-primed top strand.

In the method of the present invention, hybridized fragments are ligated. The hybridized fragments to be ligated are hybridized immediately adjacent to each other. The hybridized fragments are ligated using a suitable ligase. In one embodiment, ligation is accomplished using one or more ligases. Suitable ligases

include thermostable and non-thermostable ligases and include, but are not limited to, T4 DNA ligase, DNA ligase I, *Taq* ligase and *Tth* ligase. In another embodiment, ligation is accomplished using chemical means.

The final chimeric product can be a double-stranded chimeric polynucleotide that does not contain a contiguous, full-length template. It is therefore unnecessary to modify the template strand to facilitate its removal. This heteroduplex can be amplified using standard amplification techniques to generate homoduplex chimera or can be cloned and introduced into an organism using standard cloning and transformation techniques upon which replication *in vivo* will generate homoduplex chimeric molecules.

Chimeric polynucleotides can be selected or screened based on alterations of specific properties, *e.g.*, nucleotide structure, nucleotide function, altered enzymatic activities of proteins encoded by the chimeric polynucleotide, altered structural functions of proteins encoded by the chimeric polynucleotide, altered regulatory functions of proteins encoded by the chimeric polynucleotide, *etc.*, or a combination thereof. Subsequent selection and amplification of chimeric polynucleotides allows for the *in vitro* or *in vivo* directed evolution of biological molecules such as nucleic acid or polypeptides. This method for directed evolution would aid in the improvement of such molecules for use, for example, in medical therapies, as reagents in molecular biology, and in industry.

The present invention is particularly useful for evolving industrially or medically useful molecules for biochemical pathways, wherein the chimeric polynucleotide is itself a useful molecule (*e.g.*, promoter, aptamer, catalyst, enhancer or other regulatory element) or wherein the chimeric polynucleotide encodes a useful gene product. The chimeric polynucleotides can be or encode molecules that are more active under desired conditions to have altered or enhanced specificity, mutagenicity or fidelity. For example, desired conditions include conditions to which the reference molecule, oligonucleotide, template, or polypeptide encoded therein is not typically exposed or otherwise extreme conditions. Extreme conditions could include high or low temperature, extreme

high or low pH, extreme ionic strength, extreme solvent conditions such as organic solvent conditions, or a combination of two or more of these conditions.

Examples of industrially or medically useful polypeptides or oligonucleotides are well known in the art. Medically useful molecules include "bioactive" molecules, used herein to include peptides; proteins; polysaccharides and other sugars; lipids; and nucleic acid sequences, such as genes, and antisense molecules. Nucleic acid encoding enzymes that produce, modify or degrade polysaccharides, other sugars or lipids can be used as the scaffold, oligonucleotides or reference polynucleotide. Specific examples of bioactive molecules include, but are not limited to, insulin, erythropoietin, interferons, colony stimulating factors such as granulocyte colony stimulating factor, growth hormones such as human growth hormone, Insulin-Like Growth Factors I and II, Angiopoietin I and II, LHRH analogs, LHRH antagonists, tissue plasminogen activator, somatostatin analog, Factor VIII, Factor IX, calcitonin, dornase alpha, polysaccharides, AG337, bone inducing protein, bone morphogenic protein, brain derived growth factor, gastrin 17 immunogen, interleukins such as IL-2, PEF superoxide, permeability increasing protein-21, platelet derived growth factor, stem cell factor, thyrotropin, EGF, Tie-2 ligands, and somatomedin A and C.

One of skill in the art can readily select or design a scaffold to encode the molecule of interest to be evolved according to the method of the present invention. Methods for measuring activity of hormones, interleukins, growth factors and angiogenesis inhibitors and the like under desired conditions are well known in the art. One of ordinary skill in the art can readily determine the activity of the hormone, interleukin, growth factor or angiogenesis inhibitor encoded by the chimeric polynucleotide produced by the present invention and select those having the desired characteristics. Examples of medically useful molecules to be evolved according to the present invention also include enzymes that synthesize drugs, antibiotics, vitamins or co-factors. Other examples include vectors and genes for gene therapy. In addition, molecules that have desired therapeutic effect can be altered to lessen toxicity, antigenicity or other side effects.

Methods for determining activity under desired conditions include standard methods well known in the art. One of ordinary skill in the art can readily determine the activity of an enzyme encoded by a chimeric polynucleotide and select those oligonucleotides that encode enzymes that have desired characteristics. Enzymes include but are not limited to fermenting enzymes, proteases, lipases, oxidoreductases such as alcohol dehydrogenase, polymerases, hydrolases and luciferase.

Examples of industrially useful molecules include enzymes that synthesize polyketides, transform small molecules, hydrolyze substrates, replace steps in organic synthesis reactions or degrade pollutants such as aromatic hydrocarbons (*e.g.*, benzene, xylene, toluene and naphthalene), polychlorinated biphenyls and residual herbicides and pesticides. Catabolic pathways can be evolved using the present invention such that enzyme pathways are produced that degrade manmade pollutants that otherwise are not or only slowly catabolized. Oligonucleotides encoding such enzymes or fragments of coding regions can be used in the present invention as either the template, parent polynucleotides, a reference molecule to which chimeric polynucleotide products are compared, or combinations thereof. The method of the present invention can be used to increase, for example, the rate of an enzyme activity and the extent of the activity, *e.g.*, the affinity of the enzyme for its substrate. For example, the first enzyme in the metabolism of sulfur heterocycles by *Rhodococcus*, dibenzothiophene-monooxygenase (DBT-MO), is the bottleneck for both the rate and extent of sulfur oxidation in the biodesulfurization (BDS) process.

In one embodiment of the present invention, a chimeric polynucleotide is generated wherein one or more characteristics of the product molecule is different with respect to at least one reference polynucleotide. The difference in the chimeric polynucleotide can include a nucleotide change and/or amino acid changes in the encoded polypeptide in comparison to the reference polynucleotide, polypeptide or fragment thereof. The reference polynucleotide, polypeptide or fragment thereof can be the template or fragment, or can be a molecule related to the template used for comparison. For example, where the template is a non-functional version of a

oligonucleotide of interest or polypeptide encoded therein, then a reference molecule can be used for comparison to chimeric polynucleotides generated. The reference molecule can be a family member of the gene or gene product of interest, such as a homologous gene, or fragment thereof. One of skill in the art can readily choose a reference molecule based on the templates and oligonucleotides of interest used to generate the chimeric polynucleotides.

The characteristics to be altered according to the present invention include, but are not limited to, structural motif, stability, half-life, enzymatic activity, enzyme specificity, binding affinity, binding specificity, toxicity, antigenicity, interaction with an organism or interaction with components of an organism of the oligonucleotide or the encoded polypeptide. A functional characteristic can be altered according to the present invention such that the activity of said functional characteristic is enhanced at a higher or lower temperature compared to a reference molecule. Furthermore, said functional activities can be enhanced in various physical or chemical environments as described above or can be enhanced under standard conditions. Methods for measuring, selecting and screening these characteristics are well known in the art.

Structural motifs for proteins include, for example, α-helices, beta-sheets, solvent exposed loops, leucine zippers, β-barrel scaffolds and the like. Structural motifs for oligonucleotides include, for example, quadraplexes, aDNA, bDNA, zDNA, triple helices, stem loops, hairpins, protein binding sites and the like. Examples of regions are provided above. Methods for determining these motifs are well known in the art. In one embodiment, alteration of the characteristic includes an enhancement of the characteristic. In another embodiment, alteration of the characteristic includes a reduction in the characteristic.

In one embodiment of the present invention, a chimera is cloned prior to selection or screening. Methods of cloning oligonucleotides are well known in the art. Alternatively, the chimera can be selected or screened *in vitro* or *in vivo* prior to cloning.

The present invention allows the generation of at least one chimeric polynucleotide. The chimeric polynucleotides are different from any single

template used to generate the chimeric polynucleotide. Based on the method of the present invention, the differences can include, for example, an additional region, wherein the region is not present in the template. The additional region can be derived from an existing source of oligonucleotides, or a modified form thereof or can be a partially or completely random sequence. The additional region or regions can be present at either terminus of the resultant chimeric polynucleotide or can be present within the chimeric polynucleotide. Thus, the chimeric polynucleotide of the present invention can be longer than the template. In another embodiment, the chimeric polynucleotide can include an altered version of a region that is present in the template. The region can be the same length as the region in the hybridization template or can be longer or shorter than the region in the hybridization template. Thus, the chimeric polynucleotide can be the same size, longer or shorter than the template.

The invention will be further described with reference to the following non-limiting examples. The teachings of all the patents, patent applications and all other publications and websites cited herein are incorporated by reference in their entirety.

EXAMPLE 1

Method for Optimized Directed Evolution of PCI/TCI Polynucleotides

Heteroduplex Oligonucleotide Shuffling

Potato and tomato carboxypeptidase inhibitors (PCI and TCI, respectively) are 72 % identical at the amino acid level. To create a library of hybrid molecules from these two parents, three top strand oligonucleotides were synthesized to capture each polymorphism for the genes (Figure 2). Design modifications were carried out as described previously. Positioning of each oligonucleotide was selected to maximize the length of the perfectly base-paired interaction at the ends of each oligonucleotide without sacrificing representation of parental polymorphisms. Since no gaps were present, no polymerization was necessary and the top strand oligonucleotides were joined by ligation. DNA sequencing of 11

clones revealed between 1 and 7 crossovers per gene, with an average of 3.7 (ideal number of crossovers = 6). While each of the internal polymorphisms was represented at least once, representation of polymorphisms in the four positions nearest the junctures between the oligonucleotides were severely biased. Polymorphisms matching only the template gene were observed in 11 of 11 clones for three of these for positions and in 3 of 11 in the fourth.

The directed evolution of the PCI/TCI family of genes can be improved using synthetic oligonucleotides by optimizing the representation of allele single nucleotide polymorphisms (SNPs), dinucleotide polymorphisms (DiPs) and trinucleotide polymorphisms (TriPs) as alternative vs. degenerate loci. The mature coding regions of PCI/TCI are each 117 bp long and differ by 26 nucleotides (a 78% difference in sequence identity at the DNA level).

The PCI gene was altered to match common *E. coli* codon preferences (29 mutational changes). The TCI gene was altered in synonymous as well as non-synonymous codons. This resulted in a gene which was modified such that it contained 84% sequence identity with the original PCI gene (19 mismatches).

Mimicking *in vitro* recombination using standard degenerate oligonucleotides for these genes requires a two-fold degeneracy at each of these 19 positions, *i.e.*, to match one or the other parent, resulting in 2^{19} = 524,288-fold degeneracy. A minimum library size of over 1.5 million clones is required to capture each permutation of the parental alleles with a 95% degree of confidence. This large number is required whether a single degenerate oligonucleotide is generated or whether 19 degenerate oligonucleotides containing these 19 positions is generated. Although this number is an improvement when compared to the 2^{26}= 67 million clones which are necessary when the parents are not manipulated, further significant reductions in the required numbers would greatly increase efficiency. Focusing on the protein level, there are 11 amino acid residue differences between the two proteins. The following method of designing oligonucleotides balances the benefits of utilizing degenerate codons, *e.g.*, reduction of library size and screening, with the convenience of using commercially available synthetic methods (see Figure 2):

1. Where manipulation of parental sequences has allowed alternative codons at one locus to differ by a single nucleotide polymorphism (SNP), the alternative nucleotides at that single position are included in a two-fold degenerate locus in all oligonucleotides covering that region of the gene. The overall degeneracy of any particular oligonucleotide will be determined by the number of such SNPs and the chose termini of the oligonucleotide. These degenerate oligonucleotides will compete with alternative degenerate oligonucleotides described next. These alternative competitive oligonucleotides have identical termini.

2. Where alternative codons at a locus must differ by DiPs and TriPs, separate oligonucleotides are synthesized, each of which contain one or more of the possible permutations of the various DiPs and TriPs in the region encompassed by that oligonucleotide. For such oligonucleotides, too, the overall degeneracy is determined solely by the number of SNPs in that oligonucleotide. Since separate alternative oligonucleotides with the various permutations of DiPs and TriPs are otherwise identical, they will compete with each other for the same binding site. The termini of these oligonucleotides are identical to the desired degenerate codon oligonucleotides described above.

3. The oligonucleotides are designed to anneal perfectly at both termini to templates by synthesizing them to end in stretches of sequence identity between the two parents of, typically, 12 or more bases.

4. Other regions of the template are likewise hybridized to similarly designed degenerate and alternative degenerate oligonucleotides. Designing oligonucleotides that bind to other regions to include 5' phosphates and to abut perfectly with the neighboring oligonucleotides obviates the need for gap filling and flap trimming such that only the use of ligation is necessary to complete the chimeric strand. The need for forward and anchor oligonucleotides is also obviated, and the generation of parent clones by read-through from an upstream oligonucleotide is rendered unlikely.

For the trust of three primer binding sites, degenerate primer Deg1a is 5-fold degenerate and so consists of a mixture of 2^5=32 different primers. Since the 32 variations in Deg1b also compete for the same site, a total of 64 primers compete for this site (site 1). Likewise, there are four permutations of the four-fold degenerate Deg2 primers for a total of 16 comprising for site 2. Four permutations of 2-fold degeneracy indicate 8 primers competing for site 3. The total number of permutations of all the primers at each of the three sizes is 64 x 16 x 8=8192. Thus, the complete permutational diversity inherent in all the parental alleles can be captured in a theoretical library of 8192 clones. For 95% confidence in obtaining all of these clones, the library size (and the number of library clones screened) must be about 25,000.

EXAMPLE 2

Directed Evolution of *EGF* Gene Using TSTRAPS

Introduction

The method presented can generate every possible polymorphic permutation without bias by a protocol that involves annealing, polymerization and ligation of homoduplexed degenerate oligonucleotides. In preparation for the directed evolution of variant growth factors for differential signaling and inhibition of cellular proliferation in malignant cells, this method was applied to shuffle the genes encoding mouse and human epidermal growth factor (EGF), and to the simultaneous shuffling of EGF polymorphisms from five mammalian species. The resulting libraries of chimeric polynucleotides contained an unprecedented density of genetic crossovers and were completely free from genetic linkage. The mouse/human chimeric library represents the first gene family shuffled library to capture every possible permutation of the parental polymorphisms.

Results

Design modifications to the wild-type mouse and human EGF genes facilitates shuffling. Genes encoding the mature mouse and human EGF proteins are 74.5% identical. Modifications to these genes were made in order to allow

synthesis of optimal oligonucleotides for PARtially Scaffolded (PARSed) DNA shuffling. The design modifications include an upstream *Eco*RI site that allows for cloning of a gene encoding EGF as a fusion protein with the leader sequence of certain prokaryotic or eukaryotic expression/secretion vectors. Stop codons followed by a *Bam*HI cloning site were engineered downstream of the reading frame (Figure 3).

The design of the mouse and human *EGF* genes further included making the genes as similar as possible. This strategy required changing eleven silent polymorphisms in the mouse sequence to match the corresponding nucleotides in the human sequence. Six non-synonymous codons were also altered to reduce the polymorphic differences between them from an average of 2.5 to an average of 1, without changing the encoded amino acid residues (Figures 3 and 4). The number of nucleotide polymorphisms was thus reduced from 39 to 19, and the number of possible permutations of these clones from 239 to 219 (*i.e.*, from 5.5 x 1011 to 5 x 105 possible clones). The above manipulations reduced the total number of nucleotide permutations by six orders of magnitude without losing any of the polymorphic diversity inherent in the parental proteins.

PARSed DNA Shuffling Experimental Design

For the mouse/human *EGF* shuffling, oligonucleotides were synthesized to span the entire top strand of the modified *EGF* gene (Figure 4). Each oligonucleotide was designed to incorporate degeneracies that correspond to the polymorphisms of the mouse and human genes. In addition, polymorphic codons differing by two or three nucleotides in top strand chimeric oligonucleotides TS2 and TS3 were synthesized in separate reactions and then mixed to further reduce the degeneracy of the corresponding oligonucleotides by two-fold and four-fold, respectively. This last modification reduced the overall number of permutations needed to explore all the diversity of the wild-type parents to 6.5 x 104. Gaps of five and one nucleotide were allowed following TS1 and TS2, respectively, and thus required gap filling by DNA polymerase before ligation. TS2 and TS3 also possessed 5' phosphate groups to allow ligation. The top strand oligonucleotides

were positioned for gap filling and ligation by short bottom strand "scaffold" oligonucleotides. The scaffold oligonucleotides, however, possessed no 5' phosphate groups, and thus can not be ligated. The experimental design for the five-gene family shuffling is shown in Figure 4.

Analysis Of Mouse/Human PARSed DNA Shuffled Libraries

Products from the mouse/human PARSed DNA shuffled library were cloned. A total of 1010 chimeric genes were produced in a single PARSed shuffling reaction and a sampling of over 2 x 106 of these were captured in the first cloned library. DNA sequence analysis of random clones revealed only highly chimeric genes (Figure 5A). In 8 sequenced genes, the observed crossover density was 1 crossover per 17.5 bases, with an average of 7.75 crossovers per gene. These 8 clones also contained all 32 out of the 32 possible parental polymorphisms. Negative controls in which no polymerase or ligase was added to the PARSed DNA shuffling reaction yielded no product or clones. The distribution of polymorphisms from each parent at each polymorphic position clustered around the theoretical peak value of 50% (Figure 6). There was essentially no linkage between closely spaced parental polymorphisms. As discussed above, there are 6.5 x 104 unique permutations of the 32 polymorphisms. Since the above analysis indicates relatively little bias in generation of permutations, the number of clones needed to screen to have 99.99% probability of having screened every possible permutation in the library can be calculated. That number was calculated using the formula N = [ln(1-P)]/[ln(1-p-1)], where N is the number of screened clones, P is the probability of having screened any particular polymorphic permutation, and p is the number of possible permutations. Thus, screening 5.9 x 105 randomly chosen clones is required to screen, essentially to completion, every permutation of each parental polymorphism in these genes.

Analysis Of PARSed DNA Shuffled Libraries Of Five Mammalian Genes

EGF genes from human, mouse, rat, pig and horse differ in amino acid sequence identity by 58% to 84%. Top strand oligonucleotides were synthesized to

incorporate the polymorphisms of the parental genes and included design modifications as described above. Sequencing of 22 random clones from the chimeric library revealed crossovers between each of the 24 polymorphic positions. Seventeen of these clones are shown in Figure 5B. Single nucleotide deletions were observed in the other five clones and appear to represent artifacts within the synthesized oligonucleotide, TS5. Each of the 64 polymorphisms designed into the oligonucleotides were represented in this sampling. As was observed with the human/mouse shuffled EGF library, the frequency of crossovers between the closest alleles in these clones was the same as the frequency between the most distant alleles, and both classes centered around the ideal value of 50% (51% between the closest alleles and 50% between the most distal alleles). The number of crossovers per gene ranged from 6 to 18. The average number of crossovers in the library (11.0) differed from the theoretically perfect number of crossovers (23 crossover positions/2=11.5) by less than 5%.

Discussion

Optimal reassortment of polymorphisms in DNA shuffling is dependent on two factors. The first of these is crossover density. A typical pair of parental gene homologs that is 90% identical and only 1 kb in length will contain 100 polymorphic positions. Perfectly random recombination to explore all permutations of these polymorphisms would result in chimeric sequences averaging 50 crossovers per clone. Most other methods achieve an average of at most four crossovers for such genes. Moreover, generating multiple crossovers using current technologies becomes increasingly inefficient with decreasing gene size or increasing sequence divergence. Because of these limitations, the majority of classes of sequence permutations (*i.e.*, those involving more than a 1 crossover per 89 nucleotides (nt)) are left under-represented or entirely unexplored in the resulting chimeric libraries. The second critical parameter for optimizing recombination is the ability to achieve crossovers between close-lying polymorphisms (the ability to avoid genetic linkage effects). For hypothetical genes of 90% identity, the number of identical nucleotides between each polymorphism will average only nine bases. In the best

example reported to date, RACHITT generated 2.45 crossovers per gene between polymorphisms separated by 5 bp or fewer (Coco, W. *et al.*, *Nat. Biotechnol.* *19*:354-359, 2001). In contrast, PARSed DNA shuffling generated an average of 3.69 crossovers per gene between adjacent codons, and thus allows the testing of permutations of close-lying alleles that would otherwise tend to reassort as a single unit.

In PARSed DNA shuffling reactions, each ligation center involves three oligonucleotide participants- two top strands and a partial scaffold. Top strands that abut are ligated without polymerization. Strategically placed gaps are also used to reduce degeneracy of the annealed regions spanned by the partial scaffold. The degeneracies in the gap are introduced into the chimeric top strand during gap filling. Bottom strand oligonucleotides, *i.e.*, scaffold fragments, by contrast, are passive members in this particular embodiment of scaffolded shuffling. Bottom strand oligonucleotides can not be ligated to form a continuous strand because they do not contain, for example, a 5' phosphate. Alternatively, bottom strand oligonucleotides could be such that they can not be extended, *e.g.*, they could lack a 3' hydroxyl group. Bottom strand oligonucleotides, are not incorporated into the final library, and function only to guide homoduplex alignment of the top strands and as a source for sequence information in the small gapped regions.

The hybridizing regions of the bottom strand partial scaffolds did, in this example, contain degenerate positions. These degeneracies were designed to be perfectly complementary to the top strand chimeric oligonucleotide degeneracies at these positions. Because hybridization occurs during a gentle downward temperature ramp (*e.g.*, in this example, under conditions of high stringency), homoduplex annealing predominates over heteroduplex annealing. This encourages maximum binding strength even in regions of high sequence divergence and minimizes the required length of the scaffold, while simultaneously maximizing the specificity of binding and minimizing the representational bias of polymorphisms caused by mismatch discrimination.

Because polymorphisms are built into the degenerate oligonucleotide pools upon synthesis, physical crossovers between strands are not required. Shuffling of

the parental alleles results from ligation of any one oligonucleotide to a variety of alternative flanking oligonucleotides, as well as from polymerization across gaps in the degenerate partial scaffold oligonucleotides. Genetic linkage, a phenomenon that severely limits the sequence space explored by traditional shuffling methods, is thus absent in PARSed DNA shuffling. Recombination occurs between adjacent nucleotides as frequently as it does between distant polymorphisms. This feature allowed for the number of crossovers per gene to approach the ideal average and for the crossover density to reach 1 per 12 nt.

For the human/mouse EGF shuffling, libraries with a 1:1 ratio of the two alternative polymorphisms at each position were made. With random recombination, the libraries should have contained an ideal average number of crossovers equal to one-half of the number of potential crossover locations. This is five-fold higher than values reported for previous shuffling methods (Coco, W. *et al.*, *Nat. Biotechnol. 19*:354-359, 2001). The ideal average for the mouse/human EGF shuffling is thus 7.5 crossovers per gene. DNA sequence analysis of the PARSed DNA shuffling reaction revealed an essentially perfect average of 7.75 +/- 1.75 crossovers per gene. Similarly, the average observed for the 5-species DNA shuffled library was 11.0 +/- 2.2, which is statistically indistinguishable from the ideal number of 11.5. PARSed DNA shuffling is the first method to produce crossover densities as high as 1 per every 16 nt. It is also the first reported shuffling method that suffers no linkage effects, so that even higher crossover densities should be possible for more divergent parents. Every possible parental polymorphism in both the 2- and 5-species shuffled libraries was observed. In addition, the libraries approached the theoretical maximum of 50% reassortment at each polymorphism. These are the first gene-family DNA shuffled libraries to achieve this goal.

The unbiased linking of degenerate oligonucleotides is also important because it allows crossovers to approach the ideal distribution in short (*e.g.*, growth factor genes) or more divergent targets (as in our 5-gene library), where other multiple cross-over DNA shuffling methods become increasingly ineffective (Moore, G. *et al.*, *Proc. Natl. Acad. Sci. USA. 98*:3226-3231, 2001). To illustrate

this point, consider two close-lying alleles in two hypothetical gene homologs. Even if the chances of crossover are identical at each position along the entire length of the genes, the likelihood of crossovers between the two alleles is proportional to their separation and most unlikely for adjacent codons. In oligonucleotide based molecular breeding, the segregation of alleles can potentially be 50%, regardless of separation. This level of non-linkage, however, was not observed using oligonucleotide-based methods that rely on heteroduplex annealing, *e.g.*, RACHITT™. In contrast to the heteroduplex annealing process, the present homoduplex method allowed representation of all alleles at a frequency centered near the theoretically perfect 50%. Additionally, since each of the starting oligonucleotides contained polymorphisms from multiple parents, there was no chance of getting a significant proportion of unshuffled parental clones in the shuffled library.

Figure 4C, depicts an oligonucleotide shuffling format involving annealing of degenerate oligonucleotides to a gene-length transient template. Complex chimeric libraries can be generated in this way. A requirement for heteroduplex annealing in such methods, however limits the utility of this approach for the divergent genes used in family shuffling. Heteroduplex hybridization in divergent regions involves a compromise between polymorphism bias through mismatch discrimination under stringent annealing conditions on the one hand, and an increased proportion of non-specific products under less stringent conditions on the other. To avoid this bias, some polymorphisms must be eliminated in order to generate perfectly hybridizing anchors or "sticky feet" at the ends of each oligonucleotide. Similarly, the limitations of family shuffling by sexual PCR and other methods are well characterized. These can include generation of non-specific products, retention of unshuffled clones in the final chimeric library, severe linkage effects and, with one exception (Coco, W. *et al.*, *Nat. Biotechnol. 19*:354-359, 2001), limitation to four or fewer crossovers per gene.

Unlike other shuffling methods, PARSed DNA shuffling involves no thermocycling, stuttering, heteroduplex annealing or unmodified parental gene fragments. The single event, high stringency homoduplex hybridization shuffling

method will result in a diverse, unbiased chimeric gene library. The properties of PARSed DNA shuffling circumvent or minimize limitations of other mutagenesis or shuffling methods that rely on heteroduplex formation for gene family shuffling. The generated libraries described herein contained no observed bias, linkage or unwanted sibling and parental clones. The total number of possible permutations of the mouse/human EGF polymorphisms is 6.5 x 104. To capture 99.99% of these permutations in a random, unbiased library would require 5.9 x 105 members. Therefore, 2 x 106 chimeric *EGF* genes were cloned for this library. This is the first example of DNA shuffling that has been demonstrated to fully capture every possible parental permutation in a chimeric gene family library. The utility of design modifications that allowed for facilitated shuffling is not restricted to the examples presented here. Rather, they should be broadly applicable to any polynucleotide of interest or shuffling method. While the current application involved shuffling of small growth factor genes, it is amenable to larger sequences. Oligonucleotide-based gene synthesis protocols have been used for genes that are >1.5 kb. PARSed DNA shuffling is directly adaptable to such sizes, however for larger sequences it may be necessary to shuffle subsets of the genes that can subsequently be linked to give a full length product.

The goal of DNA shuffling is to create libraries of molecules that explore some random subset of all of the sequence space that is generated by the permutations of polymorphisms from two or more parental polynucleotides. This enormous variety of possible permutations provides a vast, diverse pool of functional protein variants from which improved protein characteristics can be selected or screened. Eliminating bias in the reassortment of polymorphisms is necessary to achieve the broadest and most representative search of the genetic diversity inherent in parental polynucleotides. The use of homoduplexed degenerate oligonucleotides to shuffle polynucleotides has achieved this goal for the genes presented herein and should be applicable to a broad range of nucleotide and protein engineering problems.

Experimental Protocols

Degenerate/alternate synthetic oligonucleotides. TS1-3 and partial scaffold (PS) 1 and 2 oligonucleotides were synthesized. Degenerate positions are indicated using IUPAC abbreviations. Otherwise identical oligonucleotides with alternative codons are distinguished by letter suffixes A and B. Oligonucleotides were synthesized by Sigma-Genosys(The Woodlands, Texas).

(SEQ ID 43) TS1A: 5'OH-
gcgcaggccggaattcagaatagtKatYctgRatgtccctYgtccYatgatgggtactgcctc

(SEQ ID 44) TS2A: 5'PO_4-
tggtgtgtgcatgYatattgaaKcattggacaagtatRcatgcaactgtgttRttggctaca

(SEQ ID 45) TS2B: 5'PO_4-
tggtgtgtgcatgYatattgaaKcattggacagctatRcatgcaactgtgttRttggctaca

(SEQ ID 46) TS3A: 5'PO_4-
cggggaKcgatgtcagtaccgagacctgaRgtggtgggaactgcgctaataggatccggctga
gcaccgcgc

(SEQ ID 47) TS3B: 5'PO_4-
cgcggggaKcgatgtcagactcgagacctgaRgtggtgggaactgcgctaataggatccggct
gagcaccgcgc

(SEQ ID 48) PS1: 5'OH- ctgacatcgMtccccgMtgtagccaaYaacacagttgcatg

(SEQ ID 49) PS2: 5'OH- ttcaatatRcatgcacacaccaYcatKgaggcagtacccatcat

Each "B" alternate oligonucleotide was combined with its "A" counterpart in equimolar amounts. The resulting five populations (TS1-3 and PS1/2) were then combined in equimolar amounts and diluted to 0.625 mM in annealing buffer.

PARSed DNA shuffling using thermophilic enzymes

Annealing was performed in 1X *Thermus aquaticus* (*Taq*) ligase buffer (NEB) supplemented with 2 mM dNTPs. The temperature was brought to 84°C for 1 minute, cooled rapidly to 75°C, ramped to 45°C over 50 minutes, and then brought rapidly to 65°C. *Taq* DNA ligase (40 U) and 0.5 U *Taq* DNA polymerase

were then added and incubated at 65°C for 40 minutes. The reaction was stopped by freezing. The resulting chimeric top strands were amplified by PCR and cloned. As a control, polymerase and ligase were omitted during the oligonucleotide assembly reactions. Subsequent PCR yielded a mixture of low molecular weight, non-specific DNA fragments. No full-length *EGF* genes were detectable upon cloning of these products.

The teachings of all references, patents and patent applications cited herein are hereby incorporated by reference in their entireties. While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

CLAIMS

What is claimed is:

1. A method for forming a chimeric polynucleotide comprising:

contacting a population of single-stranded scaffold fragments with a population of donor fragments under conditions such that at least one scaffold fragment hybridizes to at least two donor fragments at distal regions of the scaffold fragment;

treating the hybridized complexes such that single-stranded regions of the hybridized complex are filled-in; and

treating the filled-in hybridized complexes such that adjacent fragments are ligated, forming a chimeric polynucleotide.

2. The method of Claim 1, further comprising the step of trimming flaps prior to ligation.

3. The method of Claim 1, wherein the scaffold fragments comprise sequences of from about 10 to about 1000 nucleotides in length.

4. The method of Claim 1, wherein the population of scaffold fragments is derived from a single strand of a parent polynucleotide.

5. The method of Claim 1, wherein the donor fragments comprise sequences of about 10 to about 1000 nucleotides in length.

6. The method of Claim 1, wherein the donor fragments are single-stranded.

7. The method of Claim 6, wherein the population of donor fragments is derived from a single strand of a parent polynucleotide.

8. The method of Claim 1, wherein the at least one scaffold and the at least two donor fragments hybridize to each other under conditions of low stringency.

9. The method of Claim 1, wherein the population of scaffold fragments and the population of donor fragments are produced synthetically.

10. The method of Claim 1, wherein the population of scaffold fragments and the population of donor fragments are produced by cleaving a polynucleotide of interest that is a full length cDNA.

11. The method of Claim 1, wherein at least one of the fragments of the scaffold or donor populations comprises at least one region of random sequence.

12. The method of Claim 1, further comprising a step of preparing at least one single-stranded population of scaffold fragments, derived from a randomly fragmented single-stranded polynucleotide of interest.

13. The method of Claim 1, wherein the populations of scaffold and donor fragments are sufficient to form a full-length chimeric polynucleotide.

14. The method of Claim 1, further comprising screening or selecting at least one chimeric polynucleotide having desired characteristics.

15. A chimeric polynucleotide prepared according to the method of Claim 1.

16. A library of chimeric polynucleotides prepared according to the method of Claim 1.

17. The library of Claim 16, wherein the majority of the chimeric polynucleotides contain at least 3 crossover sites.

18. The library of Claim 17, wherein at least one chimeric polynucleotide contains the number of crossovers within 10% of the theoretical limit.

19. The library of Claim 18, wherein at least five chimeric polynucleotides contain the number of crossovers within 10% of the theoretical limit.

20. A method for forming at least one double-stranded chimeric polynucleotide having desired characteristics comprising:

contacting a population of scaffold fragments derived from a template polynucleotide with a population of donor fragments under conditions such that fragments of the scaffold and donor populations can hybridize to each other;

forming at least one hybridized complex comprising at least one scaffold fragment hybridized to at least two donor fragments;

treating the hybridized complex such that single-stranded regions of the hybridized complex are filled-in;

treating the filled-in hybridized complex such that adjacent fragments are ligated,

thereby forming a double-stranded chimeric polynucleotide.

21. The method of directed evolution, comprising screening or selecting at least one double-stranded chimeric polynucleotide from the library of Claim 20 having desired characteristics.

22. The method of Claim 20, further comprising trimming flaps.

23. The method of Claim 20, wherein the scaffold fragments comprise sequences that are a maximum of 25 percent as long as a polynucleotide of interest.

24. The method of Claim 20, wherein the scaffold fragments comprise sequences of from about 25 to about 1000 nucleotides in length.

25. The method of Claim 20, wherein the donor fragments comprise sequences of from about 25 to about 1000 nucleotides in length.

26. The method of Claim 20, wherein the donor fragments are single-stranded.

27. The method of Claim 26, wherein the population of donor fragments is derived from a single strand of a parent polynucleotide.

28. The method of Claim 20, wherein the scaffold and donor fragments hybridize to each other under conditions of low stringency.

29. The method of Claim 20, wherein the single-stranded regions are filled in using a polymerase.

30. The method of Claim 20, wherein the hybridized fragments are ligated using *Taq* DNA ligase or T4 DNA ligase.

31. The method of Claim 20, further comprising repeating steps hybridizing, filling in and ligating, wherein one or more chimeric polynucleotides is used to generate the populations of scaffold or donor fragments.

32. The method of Claim 20, wherein at least one of the fragments of the scaffold or donor populations comprises at least one region of random sequence.

33. A chimeric polynucleotide prepared according to the method of Claim 20.

34. A method for preparing a population of scaffold fragments, comprising the steps of:

amplifying a oligonucleotide of interest in a polymerase chain reaction, wherein the 5' terminus of a first primer comprises a 5' phosphate and wherein the 5' terminus of a second primer is devoid of a 5' phosphate;

contacting the amplified oligonucleotide with lambda exonuclease under conditions wherein oligonucleotides having a 5' phosphate are digested, leaving single-stranded oligonucleotides; and

fragmenting the single-stranded oligonucleotides, thereby preparing a population of scaffold fragments.

35. A method for forming a chimeric polynucleotide comprising:

treating a library of oligonucleotide fragments derived from a parent polynucleotide of interest and allelic variations thereof, wherein the population of fragments comprises a first population of oligonucleotides derived from one strand of the parent polynucleotide and allelic variations thereof and oligonucleotides of a second population wherein oligonucleotides are synthesized *in vitro* and derived from the other strand of the known parent polynucleotide and allelic variations thereof under conditions such that oligonucleotides of the first population can hybridize to oligonucleotides of the second population to form a gapped homoduplex;

treating the gapped homoduplex with a polymerase, wherein polynucleotide strand extension produces a double-stranded polynucleotide comprising at least one nicked strand; and

treating the nicked polynucleotide with a ligase,

thus forming a full-length polynucleotide.

36. A method of forming a single-stranded chimeric polynucleotide according to the method of Claim 35, wherein the oligonucleotides of the second population do not contain a 5' phosphate group, further comprising the step of removing the oligonucleotides of the second population after ligation.

37. The method of Claim 32, comprising the additional step of amplifying the single-stranded chimeric polynucleotide in a nucleic acid amplification reaction thereby producing more than one copy of a double-stranded chimeric polynucleotide.

38. A method of forming a single-stranded chimeric polynucleotide according to the method of Claim 35, wherein the oligonucleotides of the second population do not contain a 3' hydroxyl group, further comprising the step of removing the oligonucleotides of the second population after ligation.

39. The method of Claim 37, comprising the additional step of amplifying the single-stranded chimeric polynucleotide in a nucleic acid amplification reaction thereby producing more than one copy of a double-stranded chimeric polynucleotide.

40. The method of Claim 39, wherein the gapped homoduplex is full-length.

41. The method of Claim 35, wherein the known parent molecule sequence is from about 50 bases to about 2 kilobases in length.

42. The method of Claim 35, wherein the known parent sequence is from about 1 kilobase to about 5 kilobases in length.

43. The method of Claim 35, wherein the known parent sequence is from about 2 kilobases to about 25 kilobases in length.

44. The method of Claim 35, comprising an additional recombination step between the chimeric polynucleotide and a parent molecule or allelic variation thereof.

45. A library of chimeric polynucleotides comprising more than one chimeric polynucleotides formed according to the method of Claim 35.

46. The method of Claim 35, wherein the oligonucleotides of the second population are derived from regions of sequence identity between parent polynucleotides and allelic variations thereof.

47. The method of Claim 35, wherein the gapped homoduplex contains polymorphic sites in at least one double-stranded region of the homoduplex.

48. The method of Claim 35, wherein the gapped homoduplex contains at least one polymorphic site in the gapped region of the gapped homoduplex.

49. A method for directed evolution comprising:

forming a library of chimeric polynucleotides comprising:

contacting a first population of oligonucleotides with a second population of oligonucleotides, wherein the sequences of the first and second oligonucleotide populations are complementary to one another, under conditions such that oligonucleotides of the first population can hybridize to oligonucleotides of the second population to form a gapped homoduplex;

treating the gapped homoduplex with a polymerase, wherein polynucleotide strand extension produces a nicked polynucleotide;

treating the nicked polynucleotide with a ligase, such that nicks are ligated; and

screening the library of chimeric polynucleotides for a characteristic of interest.

50. The method of Claim 49, wherein the oligonucleotides of the first population and the oligonucleotides of the second population are derived from a known polynucleotide of interest.

51. The method of Claim 50, further comprising repeating the steps using the chimeric polynucleotide as the known polynucleotide of interest in the subsequent round of directed evolution.

52. The method of Claim 51, wherein the steps are repeated from about 2 to 50 times using a screened population of chimeric polynucleotides as the parent polynucleotides used to generate scaffold and donor fragments in a subsequent round of directed evolution.

53. The method of Claim 49, wherein the oligonucleotides of the second population do not contain 5' phosphate groups.

54. The method of Claim 49, wherein the oligonucleotides of the second population do not contain 3' hydroxyl groups.

55. The method of Claim 49, wherein the screening step comprises screening the function of the transcribed and/or translated products of the library of chimeric polynucleotides.

56. The method of Claim 49, comprising cloning the library of chimeric polynucleotides into a suitable vector prior to the screening step.

57. The method of Claim 49, further comprising:

cloning the chimeric polynucleotides into expression vectors;

transforming a suitable cell line with the cloned chimeric polynucleotides;

inducing expression of the cloned chimeric polynucleotide;

assaying the expressed product for a characteristic of interest; and

selecting the chimeric polynucleotide that expressed products with an improved characteristic of interest.

58. The method of Claim 49, further comprising:

transcribing and translating the chimeric polynucleotide *in vitro*;

assaying the transcribed and translated products for a characteristic of interest; and

selecting the chimeric polynucleotide that lead to transcribed and translated products with an improved characteristic of interest.

59. A chimeric polynucleotide formed and selected according to the method of Claim 49.

60. A method for forming a single-stranded chimeric polynucleotide comprising:

treating a library of oligonucleotide fragments derived from a parent polynucleotide of interest and allelic variations thereof, wherein the population of fragments comprises a first population of oligonucleotides derived from one strand of the parent polynucleotide and allelic variations thereof and oligonucleotides of a second population wherein oligonucleotides are synthesized *in vitro* and derived from the other strand of the known parent polynucleotide and allelic variations thereof under conditions and wherein oligonucleotides of the second population do not contain 5' phosphate groups such that oligonucleotides of the first population

can hybridize to oligonucleotides of the second population to form a gapped homoduplex;

treating the gapped homoduplex with a polymerase, wherein polynucleotide strand extension produces a double-stranded polynucleotide comprising at least one nicked strand;

treating the nicked polynucleotide with a ligase, such that the first population of oligonucleotides are ligated and the second population of oligonucleotides are not ligated; and

removing the hybridized oligonucleotides of the second population,

thus forming a single-stranded chimeric polynucleotide.

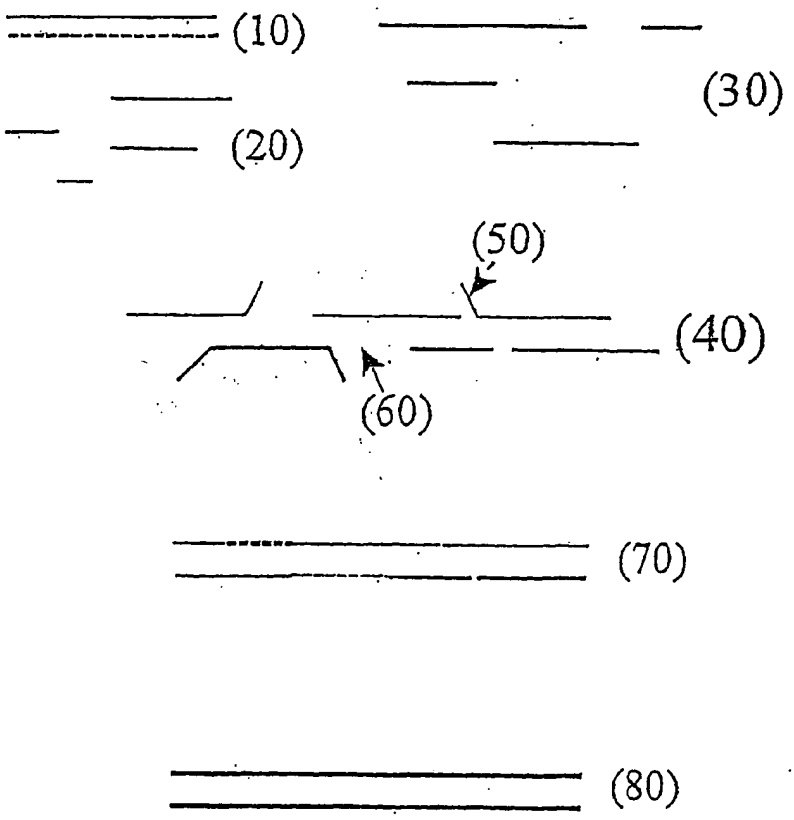


Fig. 1

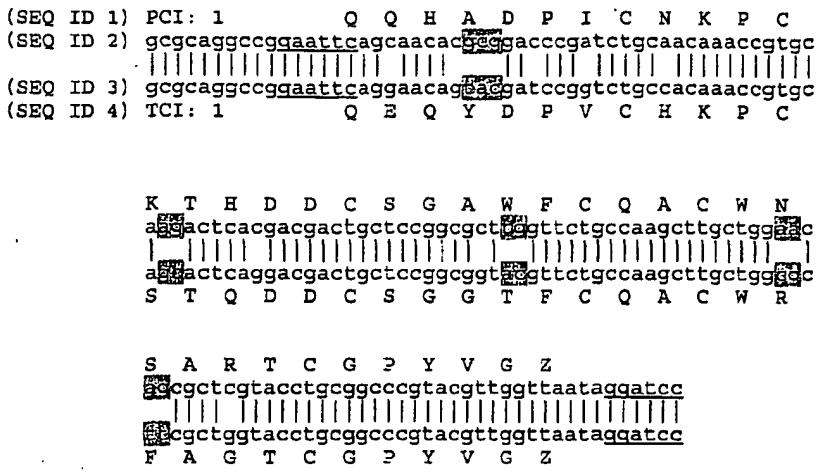

(SEQ ID 1) PCI: 1 Q Q H A D P I C N K P C
(SEQ ID 2) gcgcaggccggaattcagcaacacggcgacccgatctgcaacaaaccgtgc
(SEQ ID 3) gcgcaggccggaattcaggaacaggacgatccggtctgccacaaaccgtgc
(SEQ ID 4) TCI: 1 Q E Q Y D P V C H K P C
K T H D D C S G A W F C Q A C W N
S T Q D D C S G G T F C Q A C W R
S A R T C G P Y V G Z
F A G T C G P Y V G Z

Fig. 2A

Deg1a:
5'-phosphate -gcgcaggccggaattcag (c/g) aaca (c/g) gcgga (c/t) ccg (a/g) tctgc (a/c) acaaac
46-mer (SEQ ID NO 5)

Deg1b:
5'-phosphate -gcgcaggccggaattcag (c/g) aaca (c/g) tacga (c/t) ccg (a/g) tctgc (a/c) acaaac
46-mer (SEQ ID NO 6)

Deg2a:
5'-phosphate -cgtgcaagactca (c/g) gacgactgctccggcg (c/g) ttggttctgccaa
44-mer (SEQ ID NO 7)

Deg2b:
5'-phosphate -cgtgcaagactca (c/g) gacgactgctccggcg (c/g) tacgttctgccaa
44-mer (SEQ ID NO 8)

Deg2c:
5'-phosphate -cgtgcagcactca (c/g) gacgactgctccggcg (c/g) ttggttctgccaa
44-mer (SEQ ID NO 9)

Deg2d:
5'-phosphate -cgtgcagcactca (c/g) gacgactgctccggcg (c/g) ttacttctgccaa
44-mer (SEQ ID NO 10)

Deg3a:
5'-phosphate -gcttgctggaacagcgct (c/g) gtacctgcggcccgtacgttggttaata
47-mer (SEQ ID NO 11)

Deg3b:
5'-phosphate -gcttgctggaacttcgct (c/g) gtacctgcggcccgtacgttggttaata
47-mer (SEQ ID NO 12)

Deg3c:
5'-phosphate -gcttgctggcgcagcgct (c/g) gtacctgcggcccgtacgttggttaata
47-mer (SEQ ID NO 13)

Deg3d:
5'-phosphate -gcttgctggcgcttcgct (c/g) gtacctgcggcccgtacgttggttaata
47-mer (SEQ ID NO 14)

Fig. 2B

A)

```
                        EcoR1    >
human(SEQ ID 15)GCGCAGGCCGGAATTCAGAATAGTGACTCTGAATGTCCCCTGTCCCACGATGGGTACTGC
mouse(SEQ ID 16)-------------------------T-TC-A-G---C--ATCC--AT-T-----A------
human(SEQ ID 17)                    AsnSerAspSerGluCysProLeuSerHisAspGlyTyrCys
mouse(SEQ ID 18)                           TyrProGly      Ser   Tyr

human           CTCCATGATGGTGTGTGCATGTATATTGAAGCATTGGACAAGTATGCATGCAACTGTGTT
mouse           ---A---G---C---------C--------T--C------GC--CA--------------
human           LeuHisAspGlyValCysMetTyrIleGluAlaLeuAspLysTyrAlaCysAsnCysVal
mouse             AsnGly            His      Ser      Ser   Thr

human           GTTGGCTACATCGGGGAGCGATGTCAGTACCGAGACCTGAAGTGGTGGGAACTGCGCTAA
mouse           A-------TTCT-----T---------ACT--------ACGA--------G-----T---
human           ValGlyTyrIleGlyGluArgCysGlnTyrArgAspLeuLysTrpTrpGluLeuArgStp
mouse           Ile      Ser   Asp         Thr         Arg

                   BamH1
human           TAGGATCCGGCTGAGCACCGCGC
mouse           ------------------------
```

B)

```
                          EcoR1    >
human(SEQ ID 19)GCGCAGGCCGGAATTCAGAATAGTGATTCTGAATGTCCCTTGTCCCACGATGGGTACTGC
mouse(SEQ ID 20)-------------------------T--C---G--------C----T--------------
human(SEQ ID 21)                    AsnSerAspSerGluCysProLeuSerHisAspGlyTyrCys
mouse(SEQ ID 22)                           TyrProGly      Ser   Tyr

human           CTCCATGATGGTGTGTGCATGTATATTGAAGCATTGGACAAGTATGCATGCAACTGTGTT
mouse           ---A---G-------------C--------T---------GC---A--------------
human           LeuHisAspGlyValCysMetTyrIleGluAlaLeuAspLysTyrAlaCysAsnCysVal
mouse             AsnGly            His      Ser      Ser   Thr

human           GTTGGCTACATCGGGGAGCGATGTCAGTACCGAGACCTGAAGTGGTGGGAACTGCGCTAA
mouse           A---------C------T---------ACT----------G--------------------
human           ValGlyTyrIleGlyGluArgCysGlnTyrArgAspLeuLysTrpTrpGluLeuArgStp
mouse           Ile      Ser   Asp         Thr         Arg

                   BamH1
human           TAGGATCCGGCTGAGCACCGCGC
mouse           ------------------------
```

Fig. 3

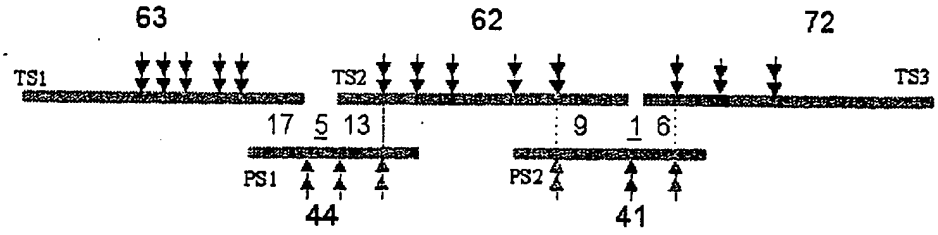


Fig. 4A

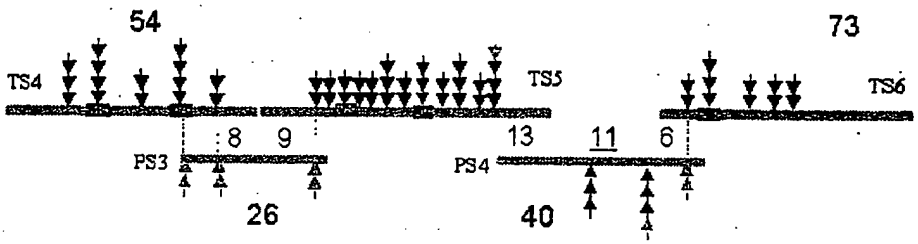


Fig. 4B

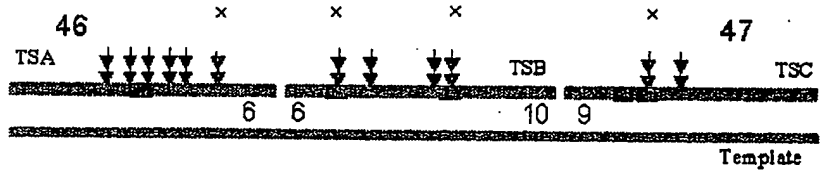


Fig. 4C

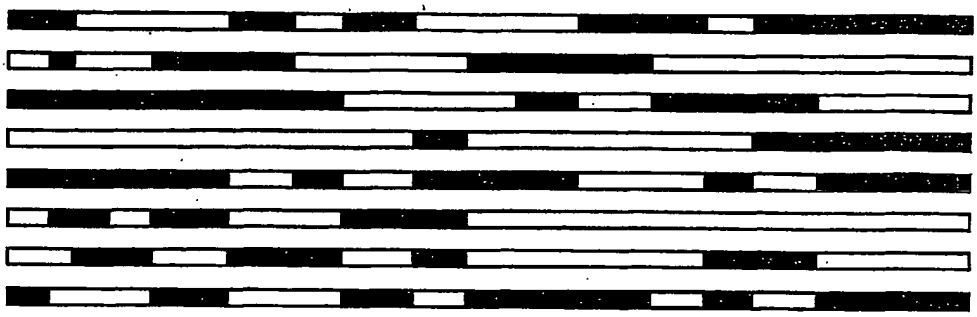


Fig. 5A

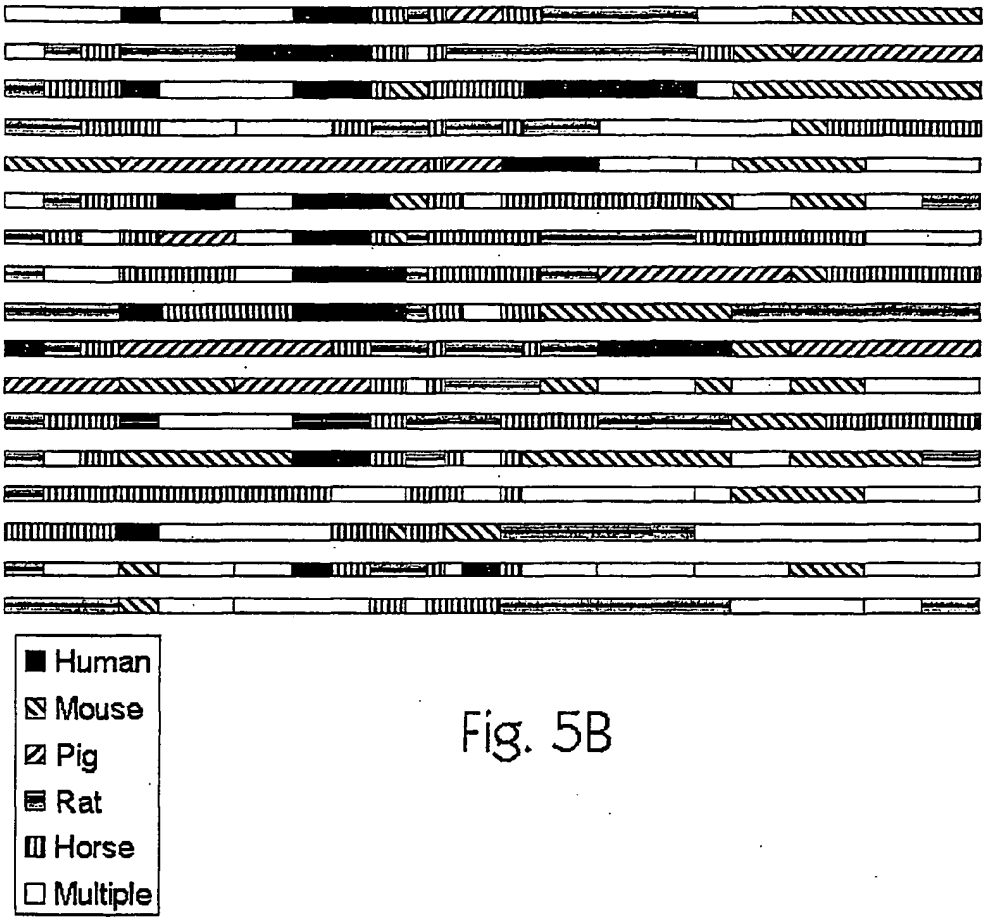


Fig. 5B

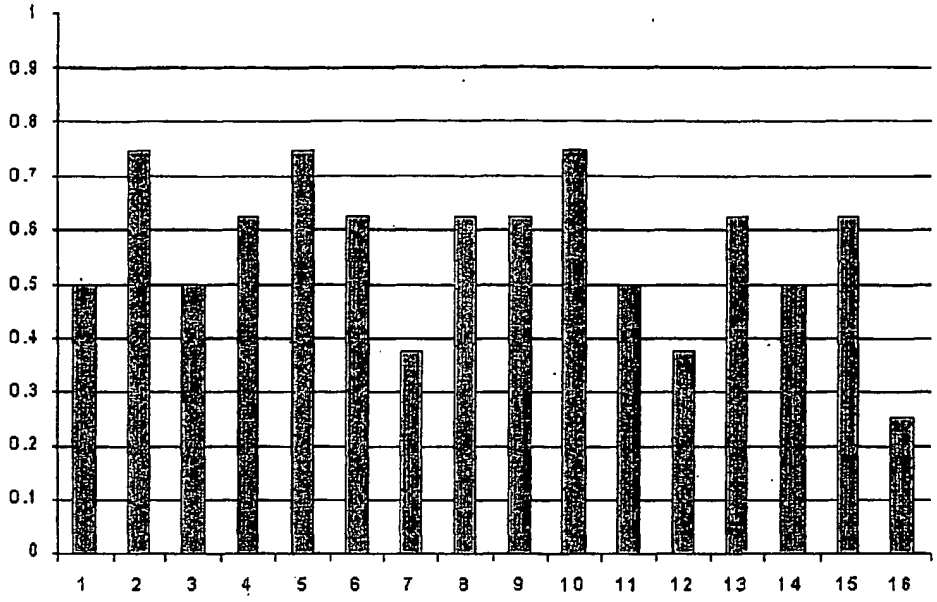


Fig. 6